

QY	1	AGTGCCTGCGGCCCTCGGGCCCTAGTACACACGACACCTGAGTGAGTGGCACACAGGAC	60
Db	1	AGTGCTTCGGCCCTCGGGCCCTAGTACACACGACCTGAGTGAGTGGCACACAGGAC	60
QY	61	CCTCTCCATGTTTAGGACCTCTCTGGGCTTCAGGAGCGTGGCGCCCGCCCTGGGCGGAC	120
Db	61	CCTCTCCATGTTTAGGACCTCTCTGGGCTTCAGGAGCGTGGCGCCCGCCCTGGGCGGAC	120
QY	121	TCCGCCCATCGCGGGCGGAATGGTCCGGGTCCGCTCCGAGTGCCTGTGCTGCTCCTCC	180
Db	121	TCCGCCCATCGCGGGCGGAATGGTCCGGGTCCGCTCCGAGTGCCTGTGCTGCTCCTCC	180
QY	181	TGGTTGCTGGGTGCAAAAGTCTGGGTTCTGGGTTTCTGGATTCCGGGCGCGTTTCAACGT	240
Db	181	TGGTTGCTGGGTGCAAAAGTCTGGGTTTCTGGATTCCGGGCGCGTTTCAACGT	240
QY	241	AGCCTGTGCGGCTCTCTCGGGTGAAGTCCGTCGCGCGCGGTGCCCGGAGCGCTAGGC	300
Db	241	AGCCTGTGCGGCTCTCTCGGGTGAAGTCCGTCGCGCGCGGTGCCCGGAGCGCTAGGC	300
QY	301	TGCGGGGGTCCGGGGCCCCAGGCAATTCGCGGTGCAGATTGACGGGATCCCGGATGCA	360
Db	258	-----CGGATGCA	265
QY	361	CCGCGGCCCTCGGCCCTCACGACGGTCCAGACCTGTGGAAAGTTCGGGAC	420
Db	266	CCGCGGCCCTCGGCCCTCACGACGGTCCAGACCTGTGTGGAAAGTTCGGGAC	325
QY	421	GGTCCCTGAGGATCCCGATGCCCTACGAGCCAGATGCTCAGCTTTATAGGTGACCTA	480
Db	326	GGTCCCTGAGGATCCCGATGCCCTACGAGCCAGATGCTCAGCTTTATAGGTGACCTA	385
QY	481	CACATGTGACTTCACTCTCAGTTTGTGATCCGTAAATGACAAATTCGAAGTACTTCA	540
Db	386	CACATGTGACTTCACTCTCAGTTTGTGATCCGTAAATGACAAATTCGAAGTACTTCA	445
QY	541	CAGTCTGTTGAGAGGATTAATGAACAATGTGTGAAAGCTTTTGCAGGAGGAGCC	600
Db	446	CAGTCTGTTGAGAGGATTAATGAACAATGTGTGAAAGCTTTTGCAGGAGGAGCC	505
QY	601	TCGGAAGCAGGCTGCGCGGACAGACACCTGCTGCACAGGACACACAGCAGCAT	660
Db	506	TCGGAAGCAGGCTGCGCGGACAGACACCTGCTGCACAGGACACACAGCAGCAT	565
QY	661	GAAGACCCCGTGGAGTGGCGGTCACTGGGATCAGACCTCGGCTTCAGCACCGCTG	720
Db	566	GAAGACCCCGTGGAGTGGCGGTCACTGGGATCAGACCTCGGCTTCAGCACCGCTG	625
QY	721	CCGAGTGGCTACCGGCTCAAGGCCAGGACGTGATGTGATGAGACTCTGTTGGCAG	780
Db	626	CCGAGTGGCTACCGGCTCAAGGCCAGGACGTGATGTGATGAGACTCTGTTGGCAG	685
QY	781	CCCAGCAGGCACCCGGCTACCCACCGACTTCGATCCGCTGGGTGGAGAGGCTAA	840
Db	686	CCCAGCAGGCACCCGGCTACCCACCGACTTCGATCCGCTGGGTGGAGAGGCTAA	745
QY	841	CAGAACAGGCGCTGGGCAAGGAGGATCGAAGGCTTGGGGCAAGGGGAGCTGTGA	900
Db	746	CAGAACAGGCGCTGGGCAAGGAGGATCGAAGTCTTGGGGCAAGGGGAGCTGTGA	805
QY	901	GACACCCCTCAAGGGGACGACCCCCACCTCACACGAGGAGGACAAATACAG	960
Db	806	GACACCCCTCAAGGGGACGACCCCCACCTCACACGAGGAGGACAAATACAG	865
QY	961	ACCATCAGCCACACCCCGTCTTACTGTGATGAGTGCCTGTTGGCTCCCGATCTGAAG	1020
Db	866	ACCATCAGCCACACCCCGTCTTACTGTGATGAGTGCCTGTTGGCTCCCGATCTGAAG	925
QY	1021	CGCCAGCTTCGGGGCCCGCGGATGGCGAAGGGGATGCGCCGAAGCTCCGTGCTCTT	1080
Db	926	CGCCAGCTTCGGGGCCCGCGGATGGCGAAGGGGATGCGCCGAAGCTCCGTGCTCTT	985

Qy	1081	GTGAGCCACCACTTACCCCAAGGGGTAGCCACTCGCCCCCGCCAGGAGGACCACT	1144
Db	986	GTGAGCCACCACTTACCCCAAGGGGTAGCCACTCGCCCCCGCCAGGAGGACCACT	1045
Qy	1141	GCAGCCATTTCACCCAGCTGGTCCCTCCAGACAGAGCGCGGGCCAGCGCAGACTCCCA	1200
Db	1046	GCAGCCATTTCACCCAGCTGGTCCCTCCAGACAGAGCGCGGGCCAGCGCAGACTCCCA	1105
Qy	1201	GAAGTTATCTATGGTGGGTATACACTCTTCACGCCCCCTTGAAGCGGGACTTTCCCATTC	1260
Db	1106	GAAGTTATCTATGGTGGGTATACACTCTTCACGCCCCCTTGAAGCGGGACTTTCCCATTC	1165
Qy	1261	CCTCACCACCTGATGTCCCGAGCACTGTGCATCCAGCCACAGTCGCCCCCCCACACAAA	1320
Db	1166	CCTCACCACCTGATGTCCCGAGCACTGTGCATCCAGCCACAGTCGCCCCCCCACACAAA	1225
Qy	1321	TGGCCCTCAGGATCTCAGGCGCTTCCACGTCAGGGGTGACCTTCGCGAGCCCTTGTTGAC	1380
Db	1226	TGGCCCTCAGGATCTCAGGCGCTTCCACGTCAGGGGTGACCTTCGCGAGCCCTTGTTGAC	1285
Qy	1381	TTCAGGGCTGCTCAGTTAGCATTTTCAGTGCCATCTACCCACAGCAGGATGGGGCCAC	1440
Db	1286	TTCAGGGCTGCTCAGTTAGCATTTTCAGTGCCATCTACCCACAGCAGGATGGGGCCAC	1345
Qy	1441	CCAGAACCAAGGCCCTTTGGAAATGATACTCTTTTCATCAGGGTTGCCTATGGGGCCAC	1500
Db	1346	CCAGAACCAAGGCCCTTTGGAAATGATACTCTTTTCATCAGGGTTGCCTATGGGGCCAC	1405
Qy	1501	GGCAGCAGGTATGCCCCCTTGCACAGGTAGAGAGCATTCATCACCCAGGGAACCCAGG	1560
Db	1406	GGCAGCAGGTATGCCCCCTTGCACAGGTAGAGAGCATTCATCACCCAGGGAACCCAGG	1465
Qy	1561	TATTAAAGAGCCCTCTGCGGGGAGACAGACATAGCAGGGTGGGCAGTGCCTCCCTTT	1620
Db	1466	TATTAAAGAGCCCTCTGCGGGGAGACAGACATAGCAGGGTGGGCAGTGCCTCCCTTT	1525
Qy	1621	ATCCTGACAACTCTAGTCGATCTTGCCCTTTTCTCCGATTGCGGATTTGGGGCCAC	1680
Db	1526	ATCCTGACAACTCTAGTCGATCTTGCCCTTTTCTCCGATTGCGGATTTGGGGCCAC	1585
Qy	1681	CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATCTCCAAATTAGTGCACACCCAGG	1740
Db	1586	CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATCTCCAAATTAGTGCACACCCAGG	1645
Qy	1741	GGCCTGGCACTCCACATCATCCATCTGCTTGTCGCAAGTGCAGATTAACCGCGTAT	1800
Db	1646	GGCCTGGCACTCCACATCATCCATCTGCTTGTCGCAAGTGCAGATTAACCGCGTAT	1705
Qy	1801	TGCC 1804	
Db	1706	TGCC 1709	
RESULT 8			
AX880429			
LOCUS	AX880429	1614 bp	DNA linear PAT 17-DEC-2003
DEFINITION	Sequence 15334 from Patent EP1074617.		
ACCESSION	AX880429		
VERSION	AX880429.1	GI:40035165	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.		
TITLE	Primers for synthesising full-length cDNA and their use		
JOURNAL	Patent: EP 1074617-A 1534 07-FEB-2001;		
FEATURES	Research Association for Biotechnology (JP)		
	Location/Qualifiers		
	1..1614		
	/organism="Homo sapiens"		
source			

```

RESULT 8
AX880429 LOCUS linear PAT 17-DEC-2003
AX880429 1614 bp DNA
DEFINITION Sequence 15334 from Patent EP1074617.
ACCESSION AX880429
VERSION AX880429.1 GI:40035165
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota, T., Iwagai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K., and Otsuki, T.;
Primers for synthesizing full-length cDNA and their use
TITLE Patent: EP 1074617-A 15334 07-FEB-2001;
JOURNAL Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..1614
/organism="Homo sapiens"
source

```

/mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 635..1168  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAE90944.1"  
 /db\_xref="GI:40035166"  
 /translation="MKTPVELAVSMQTLGHRRCGGYRVKARTSVDETLPFGSPAG  
 TRPFPDFDPWEKANRIVGKASALGANGSCETTPSRGSTPLPRKKNKVRP  
 ISHTPVSDELSFSGSESGASFGAPRVAKGDAKRALLLTPPTPRGSHSRPREAP  
 VPTQMLIRISGLPRQG"

## ORIGIN

Query Match 79.2%; Score 1434.8; DB 6; Length 1614;  
 Best Local Similarity 90.6%; Pred. No. 0;  
 Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;

```

QY 25 AGTACACAGCAGCTGAGTGGCCACAGAGACCTCTCCATGTTTAGGACCTCCT 84
DB 1 AGTACACAGCAGCTGAGTGGCCACAGAGACCTCTCCATGTTTAGGACCTCCT 60
QY 85 GGGCTCAGAGCTGGGCGCCCGCCCTGGGGGAGCTCCCCCATCCGGGGCGGAATG 144
DB 61 GGGCTCAGAGCTGGGCGCCCGCCCTGGGGGAGCTCCCCCATCCGGGGCGGATG 120
QY 145 GTCCGGGTCCGGTCCGAGTGTCTGGTGTCTCCCTGGTGTGGTGCAGAGTCTGG 204
DB 121 GTCCGGGTCCGGTCCGAGTGTCTGGTGTCTCCCTGGTGTGGTGCAGAGTCTGG 180
QY 205 GTTCGGGTCTGGATTCGGGGCCGTTTCAACGTAGCTGTGCGGGTCTTCGGGTGA 264
DB 181 GTTCGGGTCTGGATTCGGGGCCGTTTCAACGTAGCTGTGCGGGTCTTCGGGTGA 240
QY 265 GTCCGTCGGCGGGTGGCCCGGAGCGCTAGGCTGCGGGGTGCGGGGGCCCGAGG 324
DB 241 GTCCGTCGGCGGGTGGCCCGGAGCGCTAGGCTGCGGGGTGCGGGGGCCCGAGG 300
QY 325 ATTCCGGGTGCAGATTGACGGGGATCCCGGATGCACCGCGCCCGCCCTCACCG 384
DB 301 ATTCCGGGTGCAGATTGACGGGGATCCCGGATGCACCGCGCCCGCCCTCACCG 360
QY 385 ACGGTCCAGACCTGTGGGAAGAGTGGGGGACGGGTCCCTGAGGATCCGATGCCT 444
DB 361 ACGGTCCAGACCTGTGGGAAGAGTGGGGGACGGGTCCCTGAGGATCCGATGCCT 420
QY 445 ACAGCCCAAGATGCTCAGCTTTTATAGTGTGACTACACATGACTTCACTCAGTTTT 504
DB 421 ACAGCCCAAGATGCTCAGCTTTTATAGTGTGACTACACATGACTTCACTCAGTTTT 480
QY 505 GTGATCCGTAAATGAGCAAAATTCGAGACTTACAGTGTGTTGAGAGATTAATG 564
DB 481 GTGATCCGTAAATGAGCAAAATTCGAGACTTACAGTGTGTTGAGAGATTAATG 540
QY 565 AAACAATGCTGTAAAGCTCTTTGACGAGGAGGAGCTTCGGAAGCAGGGCTTGGCGGAG 624
DB 541 AAACAATGCTGTAAAGCTCTTTGACGAGGAGGAGCTTCGGAAGCAGGGCTTGGCGGAG 600
QY 625 AGCACCTCTGCTACACGGGACCAAGCAGCATGAAGACCCCGTGGAGCTGGCCGT 684
DB 601 AGCACCTCTGCTACACGGGACCAAGCAGCATGAAGACCCCGTGGAGCTGGCCGT 660
QY 685 CAGTGGATGCAGACCTGGCCCTTCAGCACCGCTCCGAGTGGTACCGGGTCAAGC 744
DB 661 CAGTGGATGCAGACCTGGCCCTTCAGCACCGCTCCGAGTGGTACCGGGTCAAGC 720
QY 745 CAGGAGCTCATATGTGATGAGACTCTGTTTGGCAGCCACAGGAGCAGCCCGCCCTACCC 804
DB 721 CAGGAGCTCATATGTGATGAGACTCTGTTTGGCAGCCACAGGAGCAGCCCGCCCTACCC 780
QY 805 ACCGGACTTCGATCCGCCCTGGGTGAGAGAGGTATACAGAACAGAGGGGTGGCAAGGA 864
DB 781 ACCGGACTTCGATCCGCCCTGGGTGAGAGAGGTATACAGAACAGAGGGGTGGCAAGGA 840
  
```

```

QY 865 GGCATCGAAGGCTTTGGGGGCAAGGGGAGCTGTGAGACCAACCCCTTAAGGGGAGCAC 924
DB 841 GGCATCGAAGGCTTTGGGGGCAAGGGGAGCTGTGAGACCAACCCCTTAAGGGGAGCAC 900
QY 925 CCCCACCTCTCACCAAGGAAGAAACAAATACAGACCCATCAGCACACCCCGTCTTA 984
DB 901 CCCCACCTCTCACCAAGGAAGAAACAAATACAGACCCATCAGCACACCCCGTCTTA 960
QY 985 CTGTGATCAGTGGCTGTTTGGCTCCCGATCTGAAGGGGCCAGCTTCGGGGGCCCGGAT 1044
DB 961 CTGTGATCAGTGGCTGTTTGGCTCCCGATCTGAAGGGGCCAGCTTCGGGGGCCCGGAT 1020
QY 1045 GCGAAGGGGATCGCCCAAGCTCCGTCTCTCTTGTGGAGCGCACCACTACCCCGAG 1104
DB 1021 GCGAAGGGGATCGCCCAAGCTCCGTCTCTCTTGTGGAGCGCACCACTACCCCGAG 1080
QY 1105 GGTAGCCACTCGCCCGCCCGCCAGGGAGCACCACTCGAGGCCATTCAACAGCTGGTCC 1164
DB 1081 GGTAGCCACTCGCCCGCCCGCCAGGGAGG----- 1109
QY 1185 CTCGAAGCAGAGCGGGGCCAGCGGAGACTCCAGAAAGTTATCTATGGTGGGTACA 1224
DB 1110 ----- 1109
QY 1225 CTCCTACGCCCCCTGAAGCGGGGACTTTCCCAATTCCTCACCCACCTGAATGTCCCAG 1284
DB 1110 ----- 1109
QY 1285 CACTGGTCAATCCAGCCACAGTCCCCCACAACAAATGGGCTCAGATCTCAGGCTTC 1344
DB 1110 -----CACCAGTCCCCCACAACAAATGGGCTCAGATCTCAGGCTTC 1154
QY 1345 CACGTACGGGTGACCTTCGGAGCCCTCGTGTACTTCCAGGGCTCGCTCAGTTAGCAT 1404
DB 1155 CACGTACGGGTGACCTTCGGAGCCCTCGTGTACTTCCAGGGCTCGCTCAGTTAGCAT 1214
QY 1405 TTCAGTGCATCTACCCCAACAGCAGGTGGGGCCACCCAGAAACCAAGGCCCTTTGGAA 1464
DB 1215 TTCAGTGCATCTACCCCAACAGCAGGTGGGGCCACCCAGAAACCAAGGCCCTTTGGAA 1274
QY 1465 ATGATCTCTTTCATCAGGTTGCTATGGGGCCACCGGACAGGTATGGCCCTTGCCA 1524
DB 1275 ATGATCTCTTTCATCAGGTTGCTATGGGGCCACCGGACAGGTATGGCCCTTGCCA 1334
QY 1525 GGTAGAGAGACATTCATCACCAGGGAACCCAGAGTATTAAGAAAGCCCTGTGGGGGC 1584
DB 1335 GGTAGAGAGACATTCATCACCAGGGAACCCAGAGTATTAAGAAAGCCCTGTGGGGGC 1394
QY 1585 AGACAGATAGCAGGGGTGGGAGTGCCTTCCCTTTATCTGTGACATCTCTAGTCGATTC 1644
DB 1395 AGACAGATAGCAGGGGTGGGAGTGCCTTCCCTTTATCTGTGACATCTCTAGTCGATTC 1454
QY 1645 TTGCTTTTCTCCCGATTGGGATTGGGGGCCACCTCTAAGATGCTCTCTCCAGCCC 1704
DB 1455 TTGCTTTTCTCCCGATTGGGATTGGGGGCCACCTCTAAGATGCTCTCTCCAGCCC 1514
QY 1705 TGCTCAACCATACTCCAAATTAGTCCAAACCCAGGGGCTGGCACTCCCAATCATCC 1764
DB 1515 TGCTCAACCATACTCCAAATTAGTCCAAACCCAGGGGCTGGCACTCCCAATCATCC 1574
QY 1765 ATTGCTCTGTCGCAAGTGCAGTAATTAACGGGTGATTGCC 1804
DB 1575 ATTGCTCTGTCGCAAGTGCAGTAATTAACGGGTGATTGCC 1614
  
```

RESULT 9  
 BD158376  
 LOCUS  
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
 ACCESSION BD158376  
 VERSION BD158376.1 GI:27864134  
 KEYWORDS JP 2002191363-A/13219.  
 SOURCE Homo sapiens (human)







QY 1165 CTCACAGACAGACGGCGGCGGAGACTCCACAGAGTTATCTATGGTGGGTACA 1224  
Db 1110 ----- 1109  
QY 1225 CTCTTCAGCCCTCGAAGCGGAGCTTCCCATTCCTCACCCACCTGAATGTCGCCAG 1284  
Db 1110 ----- 1109  
QY 1285 CACTGGTTCATCAGCACAGTCCCGCCACACAAATGGCTCAGGATCTCAGGCTTC 1344  
Db 1110 -----CACCATGTCCTCCCGCCACACAAATGGCTCAGGATCTCAGGCTTC 1154  
QY 1345 CAGTCAGGGGTGACCTTCGGAGCCCGCTGCTGACTTCCAGGGCTCGCTCAGTATGAT 1404  
Db 1155 CAGTCAGGGGTGACCTTCGGAGCCCGCTGCTGACTTCCAGGGCTCGCTCAGTATGAT 1214  
QY 1405 TTGAGTCCATCTACCCACAGAGTGGGGCCACCCAGAAACCAAGCCCTTGGAA 1464  
Db 1215 TTGAGTCCATCTACCCACAGAGTGGGGCCACCCAGAAACCAAGCCCTTGGAA 1274  
QY 1465 ATGATATCTCTTTCATCAGGGTTCCTATGGGGCCACGGCCAGGTATGGCCCTTGCCA 1524  
Db 1275 ATGATATCTCTTTCATCAGGGTTCCTATGGGGCCACGGCCAGGTATGGCCCTTGCCA 1334  
QY 1525 GGGTAGGAGACATTCATCCACCCAGGAAACCCAGGTATTAAGAGCCCTGTGGGGC 1584  
Db 1335 GGGTAGGAGACATTCATCCACCCAGGAAACCCAGGTATTAAGAGCCCTGTGGGGC 1394  
QY 1585 AGACAGATAGCAGGGGTGGGAGTGCCTCCCTTATCTGCAATCTCTAGTCATCC 1644  
Db 1395 AGACAGATAGCAGGGGTGGGAGTGCCTCCCTTATCTGCAATCTCTAGTCATCC 1454  
QY 1645 TTGCTTTTCTCCCGATTCGGATTTGGGGCCACCTCTAAGATGCTCTCTCCAGCCC 1704  
Db 1455 TTGCTTTTCTCCCGATTCGGATTTGGGGCCACCTCTAAGATGCTCTCTCCAGCCC 1514  
QY 1705 TGCTCAACATCTCAATATAGTGCACACCCAGGGCTTGCACCTCCACATCATCC 1764  
Db 1515 TGCTCAACATCTCAATATAGTGCACACCCAGGGCTTGCACCTCCACATCATCC 1574  
QY 1765 ATTGCTTGTCGCAAGTGCATATAACGGCTGATTGCC 1804  
Db 1575 ATTGCTTGTCGCAAGTGCATATAACGGCTGATTGCC 1614

RESULT 11  
LOCUS

BC022092 1407 bp mRNA linear PRI 03-OCT-2003  
DEFINITION Homo sapiens hypothetical protein FLJ14827, mRNA (cdna clone  
MGC:4570 IMAGE:3029785), complete cds.

## ACCESSION

BC022092 GI:33870282

## VERSION

BC022092.2

## KEYWORDS

MGC.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 1407)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Narusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,

Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,

Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E.,  
Schnerch, A., Schein, J.E., Jones, J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL  
MEDLINE

22388257

## PUBMED

12477932

## REFERENCE

2 (bases 1 to 1407)

## AUTHORS

Strausberg, R.

## JOURNAL

Direct Submission

Submitted (22-JAN-2002)

National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

## REMARK

COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

On Aug 19, 2003 this sequence version replaced gi:18314382.

Contact: MGC help desk

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: DCTD/DRP

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

Contact: [amadan@systemsbiology.org](mailto:amadan@systemsbiology.org)

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

## FEATURES

source

1..1407

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:4570 IMAGE:3029785"

/tissue\_type="Ovary, adenocarcinoma"

/clone\_lib="NIH MGC 9"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

1..1407

/gene="FLJ14827"

/db\_xref="LocusID:84934"

224..1033

/codon\_start=1

/product="hypothetical protein FLJ14827"

/protein\_id="AAH22092.1"

/db\_xref="GI:18314383"

/db\_xref="LocusID:84934"

/translation="MKTPELVASGMOTLGHRCRGYRVKARTSYVDETLFGSPAG

TRTPPPDFPPWVKANRTRGVGEAKGKSGCTTTPRGSTTLPTRKKNKRRP

ISHTPSYCDSEILFSGSRGAFGAPMAKGAALRLTPTPRGSHGPRPREAP

LRALHPAGPSKTEPGPAADSKLSMGLHSSRLKRLGSLTLNLVSTGHPTATSP

HTNGPQDLRPSTSGVTFFSPVTSFARSVSISVPSTPRGGATQKPKPPWK"

## ORIGIN

Query Match 75.7%; Score 1371; DB 9; Length 1407;

Best Local Similarity 100.0%; Pred. No. 2.8e-300;

Matches 1371; Conservative 0; Mismatches 0; Gaps 0;

QY 436 CCGATGCTTACAGCAAGAGTCTCAGCTTTATAGTGTGACCTACACATGACTTCAC 495

Db 1 CCGATGCTTACAGCAAGAGTCTCAGCTTTATAGTGTGACCTACACATGACTTCAC 60

QY 496 CTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTTCACAGTCTGTTGAGAG 555

Db 61 CTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTTCACAGTCTGTTGAGAG 120

QY 556 GATTAAATGAACAATGCTTTGAAAGCTCTTTGAGGAGGGAGCTTCGGAACAGGCT 615

Db 121 GATTAAATGAACAATGCTTGTAAAGCTCTTTTCAGAGGAGGCTTCGGAGCAGGGCCT 180  
 Qy 616 GGGCGGAGAGACACACCTGTGTGTACAGGAGACACAGGAGGATGAAGACCCCGGTGA 675  
 Db 181 GGGCGGAGAGACACACCTGTGTGTACAGGAGACACAGGAGGATGAAGACCCCGGTGA 240  
 Qy 676 GCTGGCGGTAGTGGGATGAGACCCCTCGGCTTTACAGCCGCTGCGGAGGTGCTACCG 735  
 Db 241 GCTGGCGGTAGTGGGATGAGACCCCTCGGCTTTACAGCCGCTGCGGAGGTGCTACCG 300  
 Qy 736 GGTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTTGGCAGCCCGCAGGACCCCG 795  
 Db 301 GGTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTTGGCAGCCCGCAGGACCCCG 360  
 Qy 796 GCCTTACCCCGGACGTCGATCCGCTCGGTGGAGAGGCTTAAACAGACCGAGCGT 855  
 Db 361 GCCTTACCCCGGACGTCGATCCGCTCGGTGGAGAGGCTTAAACAGACCGAGCGT 420  
 Qy 856 GGGCAAGGAGGATCGAAGGCTTTGGGGGCAAGGGGAGCTGTGAGACCAACCCCTCAAG 915  
 Db 421 GGGCAAGGAGGATCGAAGGCTTTGGGGGCAAGGGGAGCTGTGAGACCAACCCCTCAAG 480  
 Qy 916 GGGCAGACCCCGACCTTCACCAAGGAAGAGAACAAATACAGACCCCATGACCAAC 975  
 Db 481 GGGCAGACCCCGACCTTCACCAAGGAAGAGAACAAATACAGACCCCATGACCAAC 540  
 Qy 976 CCCGCTTACTGTGATGAGTGGCTTGGCTCCGATCTGAAGGCGGAGCTTCGGGGC 1035  
 Db 541 CCCGCTTACTGTGATGAGTGGCTTGGCTCCGATCTGAAGGCGGAGCTTCGGGGC 600  
 Qy 1036 CCCGCGATGGCAGGAGGAGTCCGCAAGTCCGTCCTTTGGAAGCCACCAAC 1095  
 Db 601 CCCGCGATGGCAGGAGGAGTCCGCAAGTCCGTCCTTTGGAAGCCACCAAC 660  
 Qy 1096 TACCCCGAGGGTACGACCTCGCCCGCCCGAGGAGGACCACTCGGAGCAATCACCC 1155  
 Db 661 TACCCCGAGGGTACGACCTCGCCCGCCCGAGGAGGACCACTCGGAGCAATCACCC 720  
 Qy 1156 AGTGTCCTTCCAGACAGAGCGGGCCAGGCGAGCTCCCAAGAGTTATCTATGGG 1215  
 Db 721 AGTGTCCTTCCAGACAGAGCGGGCCAGGCGAGCTCCCAAGAGTTATCTATGGG 780  
 Qy 1216 TGGGTTACATCTTCAGCCCTTGAAGCGGAGCTTCCATTCCTTACCCACCTGAA 1275  
 Db 781 TGGGTTACATCTTCAGCCCTTGAAGCGGAGCTTCCATTCCTTACCCACCTGAA 840  
 Qy 1276 TGTCCCGAGCACTGGTTCATCCAGCACCAGTGTGCCCCCACCACAAATGGGCTCAGGATCT 1335  
 Db 841 TGTCCCGAGCACTGGTTCATCCAGCACCAGTGTGCCCCCACCACAAATGGGCTCAGGATCT 900  
 Qy 1336 CAGGCTTCCAGGTCAGGTCACCTTCGAGAGCCCTTGTGACTTTCAGGGCTCGCTC 1395  
 Db 901 CAGGCTTCCAGGTCAGGTCACCTTCGAGAGCCCTTGTGACTTTCAGGGCTCGCTC 960  
 Qy 1396 AGTTAGCATTTAGTGCATCTACCCAGCAGGTCGGGCGCCACCAAGAACCAAGCC 1455  
 Db 961 AGTTAGCATTTAGTGCATCTACCCAGCAGGTCGGGCGCCACCAAGAACCAAGCC 1020  
 Qy 1456 CCCTTGAATATGACTCTTTCATCAGGGTGTGCTATGGGGCCACGGCAGAGTATGGC 1515  
 Db 1021 CCCTTGAATATGACTCTTTCATCAGGGTGTGCTATGGGGCCACGGCAGAGTATGGC 1080  
 Qy 1516 CCCTTGCAGGGTAGGAGGACATTCATCCAGGGAACCCAGGTTATTAAGAGAGCC 1575  
 Db 1081 CCCTTGCAGGGTAGGAGGACATTCATCCAGGGAACCCAGGTTATTAAGAGAGCC 1140  
 Qy 1576 TGTGGGGCAGACAGACATAGCAGGGGTGGGAGTGCCTCCCTTTATCTGCAATCTCT 1635  
 Db 1141 TGTGGGGCAGACAGACATAGCAGGGGTGGGAGTGCCTCCCTTTATCTGCAATCTCT 1200  
 Qy 1636 AGTCGATCTTGTCTTTTCTCCGATTCGGATTTGGGGGCGACCTTAAAGATGCTCT 1695  
 Db 1201 AGTCGATCTTGTCTTTTCTCCGATTCGGATTTGGGGGCGACCTTAAAGATGCTCT 1260

Qy 1696 CTCGAGCCCTGTCTCAACCATCTCCAAATAGTGCACCCAGGGGCTGGCACCTCCC 1755  
 Db 1261 CTCGAGCCCTGTCTCAACCATCTCCAAATAGTGCACCCAGGGGCTGGCACCTCCC 1320  
 Qy 1756 ACATCATCATCTTCTTCTGCTCCAGTGCAGATAAAGGGCGTATTGCGAA 1806  
 Db 1321 ACATCATCATCTTCTTCTGCTCCAGTGCAGATAAAGGGCGTATTGCGAA 1371

# RESULT 12 AC089999/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

AC089999 199522 bp DNA linear PRI 23-OCT-2002  
 Homo sapiens 12 BAC RP11-545P7 (Roswell Park Cancer Institute Human  
 BAC Library) complete sequence.

AC089999 AC021161  
 AC089999.20 GI:21617628

HTG.  
 Homo sapiens (human)

ORGANISM  
 Homo sapiens

REFERENCE  
 1 (bases 1 to 199522)

AUTHORS  
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbakia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
 Buhay,C., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Burch,P., Burchett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,  
 Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,  
 Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,  
 Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
 Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Hawlak,P.,  
 Haves,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,  
 Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S.,  
 Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,  
 Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U.,  
 King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,  
 Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,  
 Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X.,  
 Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,  
 Marandel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,  
 Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,  
 Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,  
 Mohabhat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,  
 Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwunonu,G.,  
 Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,  
 Rivers,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Savary,G.,  
 Scherer,S., Scott,G., Shen,H., Shim,C., Shoentari,N., Sleson,I.,  
 Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Umani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,  
 Weinstock,G. and Gibbs,R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 199522)

Worley,K.C.  
 Direct Submission

Submitted (08-FEB-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA





Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 784)  
REFERENCE Oca.T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
AUTHORS Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
TITLE Primer for synthesizing full-length cDNA and use thereof  
JOURNAL Patent: JP 2002191363-A 3935 09-JUL-2002;  
HELIX RESEARCH INSTITUTE  
COMMENT OS Homo sapiens (human)  
PN JP 2002191363-A/3935  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 2000280990  
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SATO,  
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
PI KEIICHI NAGAI, TETSUJI OTSUKI  
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC  
10,  
PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CO  
Best Local Similarity 95.7%; Pred. No.2.6e-139;  
Primer for synthesizing full-length cDNA and use thereof FH Key  
Location/Qualifiers  
FT source 1..784 /organism='Homo sapiens (human)'.  
FT Location/Qualifiers  
1..784  
/organism='Homo sapiens'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:9606'

ORIGIN

Query Match	36.5%;	Score 661.4;	DB 6;	Length 784;
Best Local Similarity	95.7%;	Pred. No.2.6e-139;		
Matches 687;	Conservative 0;	Mismatches 7;	Indels 2;	Gaps 2;

```

QY      191  GTGCAAGTCTCGGGTTCTGGGTTC TGGAATCGGGGCCGTTCCACACGTAGCCTGTGCC 250
Db              |||
QY      251  GGCTCTCTCGGTGAGTGTCGTC CGCGCGCGGTGC CC CGCGGACGGCTTAGCTGCCGGGGT 310
Db              |||
QY      311  CCGGGGCCCGGAGCATTCGGGGCTGCAGATTGACGGGGATCCCGGATGCACGCGCGCCC 370
Db              |||
QY      371  CCGGGGCCCGGAGCATTCGGGGCTGCAGATTGACGGGGATCCCGGATGCACGCGCGCCC 430
Db              |||
QY      431  GGATCCCGATCCCTACGAGCCAAAGATGCTCAGCTTTATAGTGTGACCTACATGTGAC 490
Db              |||
QY      491  TTCACTTCAGTTTTGTGATCGTAAAATGGACAATAATCGAAGCTACTTTCAGAGTCTGTT 550
Db              |||
QY      551  GAGAGGATTAATGAAACAATGCTTTGTAAGCTTTTCGAGAGGAGCCTCGGAAGCAG 610
Db              |||
QY      611  GGCCTGGCCGGAGAGCACACTGCTGTACCAGGACCAAGGAGCATGAGACCCCC 670
Db              |||
QY      671  GTGGAGCTGGCGTCAGTGGGATGAGACCCCTCGGCCCTTCAGCACCGCTGCCAGGCTGC 730
Db              |||
QY      731  TACCGGGTCAAGGCCAGGACGTCATATGTGGATG-AGACTCTGTTTGGCAGCCACGAG 789
Db              |||
QY      541  TACCGGGTCAAGGCCAGGACGTCATATGTGGATGAAGACTCTGTTTGGCAGCCACGAG 600
Db              |||

```

us-10-031-589-3.rge

[illegible]

Search completed: August 17, 2004, 08:21:00  
Job time : 7227 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 03:21:14 ; Search time 873 Seconds  
(without alignments)  
8812.700 Million cell updates/sec

Title: US-10-031-589-3  
Perfect score: 1811  
Sequence: 1 agtgcctggcgcctcggcg.....cggcgtgattgccaactgg 1811

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues 6747726  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1811	100.0	1811	4 AAF58163	Aaf58163 Human neu
2	1621	89.5	1775	4 AAH16354	Aah16354 Human CDN
3	1610.2	88.9	1826	4 ABA08585	Abao8585 Human typ
4	1434.8	79.2	1614	4 AAH16384	Aah16384 Human CDN
5	661.4	36.5	784	4 AAH07100	Aah07100 Human CDN
6	564.4	31.2	889	5 AAS75219	Aas75219 DNA encod
7	562	31.0	597	4 AAH07172	Aah07172 Human CDN
8	494	27.3	561	4 AAH12167	Aah12167 Human adu
9	436.6	24.1	477	8 ACH14455	Ach14455 Human adu
10	410.6	22.7	484	8 ACH43954	Ach43954 Human foe
11	341.8	18.9	578	4 AAH12103	Aah12103 Human CDN
12	330	18.2	440	5 AAS75218	Aas75218 DNA encod
13	138	7.6	494	9 ADB51383	Adb51383 Primaty r
14	130.8	7.2	292	6 ABN79266	Abn79266 Human ORF
15	60	3.3	60	6 AAN32144	Aan32144 Human spl
16	54	3.0	2000	7 ADA71938	Ada71938 Rice gene
17	51.6	2.8	680	4 AA199338	Aa199338 Human exc
18	51.6	2.8	680	4 AA199337	Aa199337 Human exc
19	51.6	2.8	680	5 AA163687	Aa163687 Human kid
20	51.6	2.8	680	5 AA163688	Aa163688 Human kid
21	50.8	2.8	343	4 ARI98796	Ari98796 Human exc
22	50.8	2.8	343	5 AA163192	Aa163192 Human kid
23	50.4	2.8	43981	9 ADC86790	Adc86790 Human GPC

24	50	2.8	316	6 ABL84296	Ab184296 Human ova
25	49.8	2.7	174493	7 ACA61395	Aca61395 Novel hum
26	49.8	2.7	174493	9 AAD59937	Aad59937 Human kin
27	48.4	2.7	267	6 ABL84293	Ab184293 Human ova
28	48.4	2.7	304	6 ABL84286	Ab184286 Human ova
29	47	2.6	110000	7 ABZ79565_3	Continuation (4 of
30	46.6	2.6	342	4 AAL2164	Aal12164 Human bre
31	46.6	2.6	566	4 AAL21047	Aal21047 Human bre
32	46.6	2.6	1964	4 AAF32803	Aaf32803 Human sec
33	46.2	2.6	12111	5 ABA21422	Ab21422 Human ner
34	46.2	2.6	30001	2 AAT61016	Aat61016 Total DNA
35	46.2	2.6	30001	2 AAX05110	Aax05110 S. aureof
36	45.8	2.5	172637	6 ABN83124	Abn83124 Human vol
37	45.8	2.5	349901	9 ADC86940	Adc86940 Human GPC
38	45.6	2.5	2000	7 ADA71938	Ada71938 Rice gene
39	45.6	2.5	72928	2 AAZ18355	Aaz18355 Human AST
40	45.6	2.5	72928	3 AAA80253	Aaa80253 Human AST
41	45.6	2.5	114955	2 AAX53491	Aax53491 Human ade
42	45.4	2.5	22756	4 AAL03921	Aal03921 Human rep
43	45.4	2.5	22756	5 AAS40321	Aas40321 DNA encod
44	45.2	2.5	10732	3 AAA10594	Aaa10594 Gene enco
45	45	2.5	13069	4 AAS36350	Aas36350 Human car

ALIGNMENTS

RESULT 1  
AAF58163  
ID AAF58163 standard; DNA; 1811 BP.  
AC AAF58163;  
XX  
DT 23-APR-2001 (first entry)  
DE Human neuron progenitor cell clone #2 DNA.  
XX  
KW Neuron; progenitor cell; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WC200107607-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 21-JUL-2000; 2000WO-JP004895.  
XX  
PR 23-JUL-1999; 99JP-00209817.  
PR 18-OCT-1999; 99US-0159528P.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Kawai Y;  
PI WPI; 2001-182791/18.  
XX  
PT New human polynucleotides, particularly DNAs, isolated from a cDNA  
PT library derived from progenitor cells, useful in gene therapy, as well as  
PT in producing proteins useful as diagnostic markers in drug development.  
XX  
PS Claim 1; Page 33-36; 54pp; English.  
XX  
CC The present invention relates to human proteins isolated from clones from  
CC neuron progenitor cells. The proteins and the DNA encoding them may be  
CC used in the preparation of treatments for diseases associated with the  
CC proteins  
XX  
SQ Sequence 1811 BP; 355 A; 588 C; 535 G; 333 T; 0 U; 0 Other;

Query Match 100.0%; Score 1811; DB 4; Length 1811;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 AGTGCCTCGGCGCTCTCGGCGCTAGTACACGCACTGAGTGTGGACACAGAGAC 60  
Db 1 AGTGCCTCGGCGCTCTCGGCGCTAGTACACGCACTGAGTGTGGACACAGAGAC 60  
QY 61 CCTCTCCATGTTTAGGACCTCTCTGGGCTCAGAGCGTGGCGCCCGCCCTGGCGGAC 120  
Db 61 CCTCTCCATGTTTAGGACCTCTCTGGGCTCAGAGCGTGGCGCCCGCCCTGGCGGAC 120  
QY 121 TCCGCCCATCCGCGGCGCGAATGCTCCGGTCCGCTCCGAGTGTCTGCTGCTCC 180  
Db 121 TCCGCCCATCCGCGGCGCGAATGCTCCGGTCCGCTCCGAGTGTCTGCTGCTCC 180  
QY 181 TGGTGTCTGGGTGCAAAAGTCTGGGTTCTGGATTTCGCGGCGCGTTCACACGT 240  
Db 181 TGGTGTCTGGGTGCAAAAGTCTGGGTTCTGGATTTCGCGGCGCGTTCACACGT 240  
QY 241 AGCCTGTGCCGCTCTCTGGGTGAGTCCGTCCGCGCGCGTGGCGCGGCGCTAGGC 300  
Db 241 AGCCTGTGCCGCTCTCTGGGTGAGTCCGTCCGCGCGCGTGGCGCGGCGCTAGGC 300  
QY 301 TCCGCGGGTCCGCGGCGCGAGGCAATCCGGCTGCGAGTTGACGGGATCCCGGATGCA 360  
Db 301 TCCGCGGGTCCGCGGCGCGAGGCAATCCGGCTGCGAGTTGACGGGATCCCGGATGCA 360  
QY 361 CCGCGCGCGCGCGCGCTCACCGAGCGGTCCAGACCTGGTGGGAAGAGTCCGCGGAC 420  
Db 361 CCGCGCGCGCGCGCGCTCACCGAGCGGTCCAGACCTGGTGGGAAGAGTCCGCGGAC 420  
QY 421 GGGTCCCTGAGGATCCCGATGCCCTACGAGCCAGATGCTCAGCTTTATAGTGTGACCTA 480  
Db 421 GGGTCCCTGAGGATCCCGATGCCCTACGAGCCAGATGCTCAGCTTTATAGTGTGACCTA 480  
QY 481 CACATGTGACTTCACTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTCA 540  
Db 481 CACATGTGACTTCACTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTCA 540  
QY 541 CAGTGTGTTGAGAGGATTAATGAACAAATGTTGTAAGCTCTTTGAGGAGGAGCC 600  
Db 541 CAGTGTGTTGAGAGGATTAATGAACAAATGTTGTAAGCTCTTTGAGGAGGAGCC 600  
QY 601 TCGAAGCAGGGCTCGCGCGAGACACACCTGCTCACAGGACACACAGGAGCAT 660  
Db 601 TCGAAGCAGGGCTCGCGCGAGACACACCTGCTCACAGGACACACAGGAGCAT 660  
QY 661 GAAGACCCCGTGGAGCTGCGCGTCACTGGGATGCGAGACCTCGGCTTCAGCACCGCTG 720  
Db 661 GAAGACCCCGTGGAGCTGCGCGTCACTGGGATGCGAGACCTCGGCTTCAGCACCGCTG 720  
QY 721 CCGAGTGGTACCGGGTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTGGGAG 780  
Db 721 CCGAGTGGTACCGGGTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTGGGAG 780  
QY 781 CCAGCAGGACCCCGGCTACCCACCGGACTTCGATCCGCTCGGTCGAGAGGCTAA 840  
Db 781 CCAGCAGGACCCCGGCTACCCACCGGACTTCGATCCGCTCGGTCGAGAGGCTAA 840  
QY 841 CAGAACCCAGAGGGTGGGCAAGGAGGATCGAAGGCTTTGGGGCAAGAGGAGCTGTA 900  
Db 841 CAGAACCCAGAGGGTGGGCAAGGAGGATCGAAGGCTTTGGGGCAAGAGGAGCTGTA 900  
QY 901 GACACCCCTCAAGGGGAGCACCCTCACACCAAGGAGGAGCAAAATACAG 960  
Db 901 GACACCCCTCAAGGGGAGCACCCTCACACCAAGGAGGAGCAAAATACAG 960  
QY 961 ACCCATCAGCACACCCCGTCTTACTGTGATGAGTGGTCTGGCTCCGATCGAAGG 1020  
Db 961 ACCCATCAGCACACCCCGTCTTACTGTGATGAGTGGTCTGGCTCCGATCGAAGG 1020  
QY 1021 CGCAGCTTCGGGCGCGCGGATGCGAAGGGGATGCGCAAGCTCCGTCCTCTT 1080  
Db 1021 CGCAGCTTCGGGCGCGCGGATGCGAAGGGGATGCGCAAGCTCCGTCCTCTT 1080  
QY 1081 GTGGACGCCACCACTTACCCCGAGGGGTAGCCACTCGCCCGCGCGGAGGAGGCCACCACT 1140

## RESULT 2

AAH16354

ID AAH16354 standard; cDNA; 1775 BP.

XX

AC AAH16354;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA sequence SEQ ID NO:15283.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-00116126.

XX

PR 29-JUL-1999; 99JP-00248036.





```
QY 797 CCTACCCACCGGACTTCGATCCGCTCGGCTGGAGAGGCTAACAGAACCCAGAGCGGTG 856
DB 610 CCTACCCACCGGACTTCGATCCGCTCGGCTGGAGAGGCTAACAGAACCCAGAGCGGTG 869
QY 857 GGCAGAGGAGGATCGAAGGCTTTGGGGGCAAGAGGAGCTGTGAGACCAACCCCTCAAGG 916
DB 670 GGCAGAGGAGGATCGAAGGCTTTGGGGGCAAGAGGAGCTGTGAGACCAACCCCTCAAGG 729
QY 917 GGCAGACCCCCACCCCTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 976
DB 730 GGCAGACCCCCACCCCTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 789
QY 977 CCGCTTACTGTGATGAGTCCGCTGTTTGGCTCCGATCTGAAGGCGCCAGCTTCGGGGCC 1036
DB 790 CCGCTTACTGTGATGAGTCCGCTGTTTGGCTCCGATCTGAAGGCGCCAGCTTCGGGGCC 849
QY 1037 CCGCGGATGCGAGGGGGATGCGCAAGCTCCGCTCTCTTGTGGAGCGCCACACCT 1096
DB 850 CCGCGGATGCGAGGGGGATGCGCAAGCTCCGCTCTCTTGTGGAGCGCCACACCT 909
QY 1097 ACCCCAGGGGTAGCCACTCGCCCGCCCGCCAGGAGGACCACTGGGAGCCATTACCCA 1156
DB 910 ACCCCAGGGGTAGCCACTCGCCCGCCCGCCAGGAGGACCACTGGGAGCCATTACCCA 969
QY 1157 GCTGGTCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1216
DB 970 GCTGGTCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1029
QY 1217 GGGTTACACTTTCACGCCCCCTGAAGCGGGGACTTCCCATTCCTCCACCACTGAAT 1276
DB 1030 GGGTTACACTTTCACGCCCCCTGAAGCGGGGACTTCCCATTCCTCCACCACTGAAT 1089
QY 1277 GTCCCCAGCACTGGTATCCAGGACACAGTGCCTCCCAACAAATGGGCTCAGGATCTC 1336
DB 1090 GTCCCCAGCACTGGTATCCAGGACACAGTGCCTCCCAACAAATGGGCTCAGGATCTC 1149
QY 1337 AGGCTTCCAGCTCAGGGGTGACTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1396
DB 1150 AGGCTTCCAGCTCAGGGGTGACTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1209
QY 1397 GTTAGCAATTTCAAGTCCATCTACCCACAGAGAGTGGGGCCACCCAGAAACCAAGCC 1456
DB 1210 GTTAGCAATTTCAAGTCCATCTACCCACAGAGAGTGGGGCCACCCAGAAACCAAGCC 1269
QY 1457 CTTTGGAAATGATCTTTTCATCAGGGTTCCTATGGGGCCACGGGACAGATATGGCC 1516
DB 1270 CTTTGGAAATGATCTTTTCATCAGGGTTCCTATGGGGCCACGGGACAGATATGGCC 1329
QY 1517 CTTTGGCAGGTAGAGGACATTCATCACCAGGAGACCCAGGTATTAAGAGAGCCCT 1576
DB 1330 CTTTGGCAGGTAGAGGACATTCATCACCAGGAGACCCAGGTATTAAGAGAGCCCT 1389
QY 1577 GTGGGGCAGACAGACATAGAGGGGTGGGAGTGCCTCCCTTTATCCTGACAAATCTCTA 1636
DB 1390 GTGGGGCAGACAGACATAGAGGGGTGGGAGTGCCTCCCTTTATCCTGACAAATCTCTA 1449
QY 1637 GTGATTTCTTGCCTTTTCTCCGATTCGGATTTGGGGCCACCTCTAAGATGCTCTC 1696
DB 1450 GTGATTTCTTGCCTTTTCTCCGATTCGGATTTGGGGCCACCTCTAAGATGCTCTC 1509
QY 1697 TCCAGCCCTGTCTCAACATACTCCAAATAGTGCACCAACCCAGGGGCTGGACCTCCCA 1756
DB 1510 TCCAGCCCTGTCTCAACATACTCCAAATAGTGCACCAACCCAGGGGCTGGACCTCCCA 1569
QY 1757 CATCATCATGCTTCTGCTGCAAGTCGGAATAAAGCGCGTGTATGTCACACCTGG 1811
DB 1570 CATCATCATGCTTCTGCTGCAAGTCGGAATAAAGCGCGTGTATGTCACACCTGG 1624
```

RESULT 4

AAH16384

ID AAH16384 standard; cDNA; 1614 BP.

```
XX AAH16384;
XX AC
XX XX
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA sequence SEQ ID NO:15334.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX XX
XX PD 07-FEB-2001.
XX XX
XX PF 28-JUL-2000; 2000EP-00116126.
XX XX
XX PR 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX XX
XX PA (HELI-) HELIX RES INST.
XX XX
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX XX
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX PT length cDNAs defined in the specification, and for the detection and/or
XX PT diagnosis of the abnormality of the proteins encoded by the full-length
XX PT cDNAs.
XX PS Claim 8; SEQ ID NO 15334; 2537pp + Sequence Listing; English.
XX XX
XX CC The present invention describes primer sets for synthesizing 5602 full-
XX CC length cDNAs defined in the specification. Where a primer set comprises:
XX CC (a) an oligo-dr primer and an oligonucleotide complementary to the
XX CC complementary strand of a polynucleotide which comprises one of the 5602
XX CC nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in the
XX CC specification. The primer sets can be used in antisense therapy and in
XX CC gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
XX CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX CC oligonucleotides, all of which are used in the exemplification of the
XX CC present invention
XX SQ Sequence 1614 BP; 317 A; 514 C; 487 G; 296 T; 0 U; 0 Other;
```

Query Match 79.2%; Score 1434.8; DB 4; Length 1614;  
Best Local Similarity 90.6%; Pred. No. 0;  
Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;

QY 25 AGTACACACGACCTGAGTGTGGACACGAGGACCTCTCCATGTTTAGGACCTCTCT 84

Db 1 AGTACACACGACCTGAGTGTGGACACGAGGACCTCTCTCCATGTTTAGGACCTCTCT 60

QY 85 GGGCCTCAGGAGCGTGGCGCCCGCCCTGGGGGAGCTCCCCCATCCCGGCGCGAATG 144

Db 61 GGGCCTCAGGAGCGTGGCGCCCGCCCTGGGGGAGCTCCCCCATCCCGGCGCGGATG 120

QY 145 GTCCGGGTCCGGTCCGAGTGTCTGCTGGCTGCTCCCTGGTGTGGTGCAGAGTCTGG 204  
 DB |||||  
 QY 121 GTCCGGGCGCGTCCGAGTGTCTGCTGGCTGCTCCCTGGTGTGGTGCAGAGTCTGG 180  
 DB |||||  
 QY 205 GTTCTGGGTCTTGATTCGGGGCGGTTTACAGTGTAGCTGTGGCGGCTCTCGAGTGA 264  
 DB |||||  
 QY 181 GTTCTGGGTCTTGATTCGGGGCGGTTTACAGTGTAGCTGTGGCGGCTCTCGAGTGA 240  
 DB |||||  
 QY 265 GTCCGTTCGGCGCGGTCGCCCGGACGGCTTAGCTGTCCGGGGTCCGGGGCCCCAGGC 324  
 DB |||||  
 QY 241 GTCCGTTCGGCGCGGTCGCCCGGACGGCTTAGCTGTCCGGGGTCCGGGGCCCCAGGC 300  
 DB |||||  
 QY 325 ATTCCGGGCTGCAGATTGACGGGGATCCGGGATCCGGATGACCGCGCGCGCCCTCACCG 384  
 DB |||||  
 QY 301 ATTCCGGGCTGCAGATTGACGGGGATCCGGGATGACCGCGCGCGCCCTCACCG 360  
 DB |||||  
 QY 385 ACCGGTCCAGACCTGTGTGGGAAGGTGCGGGGACGGGTCCCTGAGGATCCCGATGCCCT 444  
 DB |||||  
 QY 361 ACCGGTCCAGACCTGTGTGGGAAGGTGCGGGGACGGGTCCCTGAGGATCCCGATGCCCT 420  
 DB |||||  
 QY 445 ACAGACCAAGATGCTAGCTTTTATAGTGTGACTTACATGTGACTTACCTCACTTTT 504  
 DB |||||  
 QY 421 ACAGACCAAGATGCTAGCTTTTATAGTGTGACTTACATGTGACTTACCTCACTTTT 480  
 DB |||||  
 QY 505 GTGATCCGTAAATGGACAATTCGAGTACTTACAGTGTCTGTGAGAGGATTAATG 564  
 DB |||||  
 QY 481 GTGATCCGTAAATGGACAATTCGAGTACTTACAGTGTCTGTGAGAGGATTAATG 540  
 DB |||||  
 QY 565 AAACAATGCTTTGTAAGCTCTTTGAGGAGGAGCTTCGGAAGCAGGGCTTGGCCGGCAG 624  
 DB |||||  
 QY 541 AAACAATGCTTTGTAAGCTCTTTGAGGAGGAGCTTCGGAAGCAGGGCTTGGCCGGCAG 600  
 DB |||||  
 QY 625 AGCACACTGCTGTCACAGGACCAAGAGGATGAGACCCCGTGGAGTGGCGCT 684  
 DB |||||  
 QY 601 AGCACACTGCTGTCACAGGACCAAGAGGATGAGACCCCGTGGAGTGGCGCT 660  
 DB |||||  
 QY 685 CAGTGGGATGAGACCTTCGAGCAGGCTGCGGAGTGGCTACCGGTCAAGGC 744  
 DB |||||  
 QY 661 CAGTGGGATGAGACCTTCGAGCAGGCTGCGGAGTGGCTACCGGTCAAGGC 720  
 DB |||||  
 QY 745 CAGGACGTATATGTGATGAGACTCTGTTGGAGCCGAGGACACCGGCTACCCC 804  
 DB |||||  
 QY 721 CAGGACGTATATGTGATGAGACTCTGTTGGAGCCGAGGACACCGGCTACCCC 780  
 DB |||||  
 QY 805 ACCGACTTCGATCCGCTGGGTGGAGAGGCTTAACAGAACAGAGGCTGGGCAAGCA 864  
 DB |||||  
 QY 781 ACCGACTTCGATCCGCTGGGTGGAGAGGCTTAACAGAACAGAGGCTGGGCAAGCA 840  
 DB |||||  
 QY 865 GGATCGAAGGCTTTGGGGCAAGGGGAGCTGTGAGACACCCCTCAAGGGGCGAGC 924  
 DB |||||  
 QY 841 GGATCGAAGGCTTTGGGGCAAGGGGAGCTGTGAGACACCCCTCAAGGGGCGAGC 900  
 DB |||||  
 QY 925 CCCACCTCAGACCAAGGAGAGACAAATACAGACCCATCAGACACACCCGCTTA 984  
 DB |||||  
 QY 901 CCCACCTCAGACCAAGGAGAGACAAATACAGACCCATCAGACACACCCGCTTA 960  
 DB |||||  
 QY 985 CTGTGATGAGTCTGCTGTTGGCTCCCGATCTGAAGCGCCAGCTTCGGGGCCCCGGAT 1044  
 DB |||||  
 QY 961 CTGTGATGAGTCTGCTGTTGGCTCCCGATCTGAAGCGCCAGCTTCGGGGCCCCGGAT 1020  
 DB |||||  
 QY 1045 GGGAAAGGGGATCCCGAAGCTTCGCTCTGTTGGAGCCGACCCACCTACCCCGAG 1104  
 DB |||||  
 QY 1021 GGGAAAGGGGATCCCGAAGCTTCGCTCTGTTGGAGCCGACCCACCTACCCCGAG 1080  
 DB |||||  
 QY 1105 GGGTAGCCTTCGCGCCCGCCCGAGGAGGACCACTGCGAGCCATTCAACCCAGCTGGTC 1164  
 DB |||||  
 QY 1081 GGGTAGCCTTCGCGCCCGCCCGAGGAGG----- 1109  
 DB |||||  
 QY 1165 CTCCAAGACAGACCGGGGCGAGCGGAGACTCCAGAGTTATCTATGGTGGGTACA 1224  
 DB |||||  
 QY 1110 ----- 1109

QY 1225 CTCTTCACGCCCCCTGAAGCGGGACTTTCCTTCCCTCACCACCTGATGTCCCCAG 1284  
 DB |||||  
 QY 1285 CACTGGTTCATCAGAGCCACAGTGGCCCCCACCACAAATGGGCTCAGGATCTCAGGCTTC 1344  
 DB |||||  
 QY 1110 -----CACCAGTGGCCCCCACCACAAATGGGCTCAGGATCTCAGGCTTC 1154  
 DB |||||  
 QY 1345 CAGTCAAGGAGTACCTTCGAGAGCCCTTGTGTGACTTCAAGGCTCGCTCAGTTAGCAT 1404  
 DB |||||  
 QY 1155 CAGTCAAGGAGTACCTTCGAGAGCCCTTGTGTGACTTCCAGGCTCGCTCAGTTAGCAT 1214  
 DB |||||  
 QY 1405 TTCAGTGCATCTACCCACAGCAGGTGGGGCCACCAGAAACCAAGCCCTTCGAA 1464  
 DB |||||  
 QY 1215 TTCAGTGCATCTACCCACAGCAGGTGGGGCCACCAGAAACCAAGCCCTTCGAA 1274  
 DB |||||  
 QY 1465 ATGATACCTTTTCATCAGGTTGCTATGGGGCCACGGGACAGGTATGCGCCCTTCGCA 1524  
 DB |||||  
 QY 1275 ATGATACCTTTTCATCAGGTTGCTATGGGGCCACGGGACAGGTATGCGCCCTTCGCA 1334  
 DB |||||  
 QY 1525 GGTAGAGGACATTCATCACCAGGAGCCAGGATTTAAAGAGCCCTGTGGGGC 1584  
 DB |||||  
 QY 1335 GGTAGAGGACATTCATCACCAGGAGCCAGGATTTAAAGAGCCCTGTGGGGC 1394  
 DB |||||  
 QY 1585 AGACAGACATAGCAGGGGTGGGACAGTGCCTCTCTGCAATCTCTAGTCGATTC 1644  
 DB |||||  
 QY 1395 AGACAGACATAGCAGGGGTGGGACAGTGCCTCTCTGCAATCTCTAGTCGATTC 1454  
 DB |||||  
 QY 1645 TTGCTTTTTCCTCCGATTCGGATTTGGGGGCCACCTCTAAGATGCTCTCTCCAGCC 1704  
 DB |||||  
 QY 1455 TTGCTTTTTCCTCCGATTCGGATTTGGGGGCCACCTCTAAGATGCTCTCTCCAGCC 1514  
 DB |||||  
 QY 1705 TGCTCAACCACTATCCAAATTAGTGCCACCCAGGGGCTGGCACCTCCACATCATCC 1764  
 DB |||||  
 QY 1515 TGCTCAACCACTATCCAAATTAGTGCCACCCAGGGGCTGGCACCTCCACATCATCC 1574  
 DB |||||  
 QY 1765 ATTGCTTTGTCGCAAGTGCAGATTAACCGGCTGATTGCC 1804  
 DB |||||  
 QY 1575 ATTGCTTTGTCGCAAGTGCAGATTAACCGGCTGATTGCC 1614  
 DB |||||

## RESULT 5

AAH07100  
 ID AAH07100 standard; cDNA; 784 BP.  
 XX  
 AC AAH07100;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA clone (5'-primer) SEQ ID NO:3935.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

PS Claim 1; SEQ ID NO 3935; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 784 BP; 174 A; 224 C; 251 G; 130 T; 0 U; 5 Other;

Query Match 36.5%; Score 661.4; DB 4; Length 784;  
Best Local Similarity 98.7%; Pred. No. 1.2e-168;  
Matches 687; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 191 GTGCAAGTCTGGTTCTGGTTCTGGATTTCGGCGCGCGTTCACACGTAGCTGTGCC 250  
DB 1 GTGCAAGTCTGGTTCTGGTTCTGGATTTCGGCGCGCGTTCACACGTAGCTGTGCC 60  
QY 251 GGCTCTCTCGGTGAGTCCGCGCGCGTTCGGCGCGCGTTCGGCGCGCGTTCGGCGGT 310  
DB 61 GGCTCTCTCGGTGAGTCCGCGCGCGTTCGGCGCGCGTTCGGCGCGCGTTCGGCGGT 120  
QY 311 CCGGGGCGCCAGGATTCGGGGTTCGAGATTCGAGGGATCCGGATGACACCGCGGCC 370  
DB 121 CCGGGGCGCCAGGATTCGGGGTTCGAGATTCGAGGGATCCGGATGACACCGCGGCC 180  
QY 371 CCGGGGCGCTCAGCAGCGGTCCAGACCTGGTGGGAAGAGTGGCGGCGGTCCCTGA 430  
DB 181 CCGGGGCGCTCAGCAGCGGTCCAGACCTGGTGGGAAGAGTGGCGGCGGTCCCTGA 240  
QY 431 GGATCCCGATGCTTACGAGCAAGATCTCAGCTTTTATAGGTGACCTACATGTGAC 490  
DB 241 GGATCCCGATGCTTACGAGCAAGATCTCAGCTTTTATAGGTGACCTACATGTGAC 300  
QY 491 TTCACTTCAGTTTGTGATCCGTAATGGACAAATTCGAAGTACTTTCACATGCTGTT 550  
DB 301 TTCACTTCAGTTTGTGATCCGTAATGGACAAATTCGAAGTACTTTCACATGCTGTT 360  
QY 551 GAGAGGATTAATGAACAATGCTTTGAAAGCTTTTCAGAGGGAGCGCTCGGAAGCAG 610  
DB 361 GAGAGGATTAATGAACAATGCTTTGAAAGCTTTTCAGAGGGAGCGCTCGGAAGCAG 420  
QY 611 GGCTTGGCGGACAGACACCTGCTTCACAGGACACAGGACGATGAGACCC 670  
DB 421 GGCTTGGCGGACAGACACCTGCTTCACAGGACACAGGACGATGAGACCC 480  
QY 671 GTGAGGTGGCGGTCAGTGGGATGCAGACCTCGGCTTCAGACCGCTCCGAGGTGGC 730  
DB 481 GTGAGGTGGCGGTCAGTGGGATGCAGACCTCGGCTTCAGACCGCTCCGAGGTGGC 540

QY 731 TACCGGGTCAAGCCAGGACGTCATATGTGGATG-AGACTCTGTTGGCAGCCAGCAGG 789  
DB 541 TACCGGGTCAAGCCAGGACGTCATATGTGGATGAGACTCTGTTGGCAGCCAGCAGC 600  
QY 790 CACCGGGCTACCCACCGGACTTCGATCCCGCTGGTGGAGAGGCTAACAGACCAG 849  
DB 601 CACCGGGCTACCCACCGGACTTCGATCCCGCTGGTGGAGAGGCTAACAGACCAG 660  
QY 850 AGGCGTGGCGCAAGGA-GGCATCGAAGGCGCTTGGGG 884  
DB 661 AGGCGTGGCGCAAGGANGGATCCAGGCGCTTGGGG 696

#### RESULT 6

AA575219  
ID AAS75219 standard; cDNA; 889 BP.

XX AAS75219;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #11023.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG11032.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX PS Claim 1; SEQ ID NO 11023; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
sequences. (I) is useful as hybridisation probes, polymerase chain  
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
and in recombinant production of (II). The polynucleotides are also used  
in diagnostics as expressed sequence tags for identifying expressed  
genes. (I) is useful in gene therapy techniques to restore normal  
activity of (II) or to treat disease states involving (II). (II) is  
useful for generating antibodies against it, detecting or quantitating a  
polypeptide in tissue, as molecular weight markers and as a food  
supplement. (II) and its binding partners are useful in medical imaging  
of sites expressing (II). (I) and (II) are useful for treating disorders  
involving aberrant protein expression or biological activity. The  
polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AA594564 represent novel human diagnostic  
coding sequences of the invention. Note: The sequence data for this  
patent did not appear in the printed specification, but was obtained in  
electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences

```
SQ Sequence 889 BP; 187 A; 265 C; 246 G; 191 T; 0 U; 0 Other;
Query Match 31.2%; Score 564.4; DB 5; Length 889;
Best Local Similarity 97.2%; Pred. No. 2.3e-142;
Matches 585; Conservative 0; Mismatches 16; Indels 1; Gaps 1;
QY 13 CCTCGGGGCTAGTACACAGCAGCTGAGTGGACAGAGACCTCTCCATGTT 72
D 1 CCTCGGGGCTAGTACACAGCAGCTGAGTGGACAGAGACCTCTCCATGTT 60
QY 73 TAGGAGCTCTCGGGCTCAGGAGCGTGGCGCCCGCCCTGGGCGGACTCCCCCATCCG 132
D 61 TAGGAGCTCTCGGGCTCAGGAGCGTGGCGCCCGCCCTGGGCGGACTCCCCCATCCG 120
QY 133 CGGCGGGAATGCTCGGGTGGCGTCCGAGTGTCTGCTGCTCTCCCTGGTGGGT 192
D 121 CGGCGGGAATGCTCGGGGCGGCTCCGAGTGTCTGCTGCTCTCCCTGGTGGGT 180
QY 193 GCAGAGTCTGGGTTC-TGGTCTTCTGATTTCGGGCGCTTCACAGTAGCTGTGCCG 251
D 181 GCAGAGTCTGGGTTC-TGGTCTTCTGATTTCGGGCGCTTCACAGTAGCTGTGCCG 240
QY 252 GCTCCTCGGGTGAAGTCCGTCGGCGCGGTGCCCGCGGACGGCTAGGTCGCGGGT 311
D 241 GCTCCTCGGGTGAAGTCCGTCGGCGCGGTGCCCGCGGACGGCTAGGTCGCGGGT 300
QY 312 CGGGGCCCCAGGCAATCCGGGCTGCAGATTGACGGGATCCCGGATGACCGCGCCCC 371
D 301 CGGGGCCCCAGGCAATCCAGCTGCATATTGACAGGATCCCGGATGACCGCGCCCC 360
QY 372 CGCGCCCTCCAGGAGCTCCAGACCTGTGGTGGGAAGTGGCGGAGCGGTCCTGAG 431
D 361 CGCGCCCTCCAGGAGCTCCAGACCTGTGGTGGGAAGTGGCGGAGCGGTCCTGAG 420
QY 432 GATCCGATGCTACGAGCAAGATGCTCAGCTTTATAGTGTGACCTACATGTGACT 491
D 421 GATCCGATGCTACGAGCAAGATGCTCAGCTTTATAGTGTGACCTACATGTGACT 480
QY 492 TCACCTCAGTTTGTGATCCGTAATGGAATTCGAAATTCGAAATTCGAAATTCGAA 551
D 481 TCACCTCAGTTTGTGATCCGTAATGGAATTCGAAATTCGAAATTCGAAATTCGAA 540
QY 552 AGAGGATTAATGAACAAATGCTGTGTAAGCTCTTTGAGGAGGAGCTTCGGAACAG 611
D 541 AGAGGATTAATGAACAAATGCTGTGTAAGCTCTTTGAGGAGGAGCTTCGGAACAG 600
QY 612 GC 613
D 601 GC 602
```

## RESULT 7

AAH07172

AAH07172 standard; cDNA; 597 BP.

AC AAH07172;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:4007.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

```
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 1; SEQ ID NO 4007; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification, where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the 5602
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention.
XX
SQ Sequence 597 BP; 102 A; 179 C; 193 G; 117 T; 0 U; 6 Other;
Query Match 31.0%; Score 562; DB 4; Length 597;
Best Local Similarity 97.1%; Pred. No. 8.8e-142;
Matches 579; Conservative 0; Mismatches 16; Indels 1; Gaps 1;
QY 25 AGTACACAGCAGCTGAGTGGTGGACAGAGACCTCTCCATGTTTAGGACCTCT 84
D 1 AGTACACAGCAGCTGAGTGGTGGACAGAGACCTCTCTCCATGTTTAGGACCTCT 60
QY 85 GGGGCTCAGGAGCGTGGCGCCCGCCCTGGGCGGACTCCCGCCCATCCGCGGGCGGAATG 144
D 61 GGGGCTCAGGAGCGTGGCGCCCGCCCTGGGCGGACTCCCGCCCATCCGCGGGCGGATG 120
QY 145 GTCCGGTCCGCTCCGAGTGTCTGCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTG 204
D 121 GTCCGGTCCGCTCCGAGTGTCTGCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTG 180
QY 205 GTTCTGGGTTTCTGGATTTCGGGCGCGCTTCACAGTAGCTGTGCCGGCTCTCGGGTGA 264
D 181 GTTCTGGGTTTCTGGATTTCGGGCGCGCTTCACAGTAGCTGTGCCGGCTCTCGGGTGA 240
QY 265 GTCCGTCGCGCGCGGTGCGCGGACGGGCTAGGCTGCGCGGGGTCCGGGGGCCCGAGCC 324
D 241 GTCCGTCGCGCGCGGTGCGCGGACGGGCTAGGCTGCGCGGGGTCCGGGGGCCCGAGCC 300
QY 325 ATTCCGGGCTGCAGATTGACGGGATCCCGGATGACCGCGCGCCCGCGGCTCTCACCG 384
D 301 ATTCCGGGCTGCAGATTGACGGGATCCCGGATGACCGCGCGCCCGCGGCTCTCACCG 360
QY 385 ACGGTCACAGCTGTGGGAAGAGTGGGGAACGGGTCCCTGAGGATCCCGATGCTCT 444
```



Db	361	ACGGGTCCAGACCTGGTGGGAAGAAAGGTGCGNGGACGGGTCCCTGAGGATCCCNATGCCT	420
Qy	445	ACGAGCCAAAGATGCTCAGCTTTATAGGTGTGACCTACACATGTGACCTTCACTCAGTTTT	504
Db	421	ACGAGCCAAAGATGCTCAGCTTTATAGGTGTGACCTACACATGTGACCTTCACTCAGTTTT	480
Qy	505	GTGATCCGTAATAATGGACAAATTCGAAGCTACTTTCAC-AGTGTGTTGAGAGGATTAAAT	563
Db	481	GTGATCCGTAATAATGGACAAATTCGAAGCTACTTTCACAGTGTGTTGATAGGATTAAAT	540
Qy	564	GAACAATGCTTGTAAAGCTCTTTTCAGGAGGAGCGCTTCGGAAGCAGGGCTCGCC	619
Db	541	GAACAATGCTTGTAAAGCTCTTTTCAGGAGGAGCGCTTCGGAAGCAGGGCTCGCC	596
RESULT 8			
AAH12167/c			
ID	AAH12167 standard; cDNA; 561 BP.		
XX	AAH12167;		
AC	XX		
XX	XX		
DT	26-JUN-2001 (first entry)		
XX	XX		
DE	Human cDNA clone (3'-primer) SEQ ID NO:9002.		
XX	XX		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
OS	Homo sapiens.		
XX	XX		
PN	EP1074617-A2.		
XX	XX		
PD	07-FEB-2001.		
XX	XX		
PF	28-JUL-2000; 2000EP-00116126.		
XX	XX		
PR	29-JUL-1999; 99JP-00248036.		
PR	27-AUG-1999; 99JP-00300253.		
PR	11-JAN-2000; 2000JP-00118776.		
PR	02-MAY-2000; 2000JP-00183767.		
PR	09-JUN-2000; 2000JP-00241899.		
XX	XX		
PA	(HELI-) HELIX RES INST.		
XX	XX		
PI	Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Osuki T;		
XX	XX		
DR	WPI; 2001-318749/34.		

CC	AAH13633	to AAH18742	represent human cDNA sequences; AAB92446 to AAB95893
CC	represent human amino acid sequences; and AAH13629 to AAH13632	represent	
CC	oligonucleotides, all of which are used in the exemplification of the		
CC	present invention		
XX			
SQ	Sequence 561 BP; 118 A; 142 C; 181 G; 117 T; 0 U; 3 Other;		
Query Match			
Best Local Similarity 99.4%; Score 494; DB 4; Length 561;			
Matches 494; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1308	CCCCCACACAAATGGGCGCTCAGGATCTCAGGCGCTTCCACGTCAGGGGTCACCTTCGCGA	1367
DB	497	CCCCCCACANAATGGGCGCTCAGATNTCAGGCGCTTCCACGTCAGGGTCACCTTCGCGA	438
QY	1368	GGCCCGCTGGTGACTTCACGGGCTCGCTCAGTTAGCATTTTCAGTGCCATCTACCCACGAC	1427
DB	437	GGCCCGCTGGTGACTTTCAGGGCTCGCTCAGTTAGCATTTTCAGTGCCATCTACCCACGAC	378
QY	1428	GAGGTGGGGCCACCCAGAAACCAAGCCCGCTTGGAAATGATATCTTTTCATCAGGGTTG	1487
DB	377	GAGGTGGGGCCACCCAGAAACCAAGCCCGCTTGGAAATGATATCTTTTCATCAGGGTTG	318
QY	1488	CCTATGGGGCCACGGGACAGGTATGGCCCGCTTCCACGGGTAGGAGACATTCATCACCC	1547
DB	317	CCTATGGGGCCACGGGACAGGTATGGCCCGCTTCCACGGGTAGGAGACATTCATCACCC	258
QY	1548	AGGGAACCCAGGTATTAAAGAGCCCGCTGTGGGGCAGACACATAGCAGGGTGGGC	1607
DB	257	AGGGAACCCAGGTATTAAAGAGCCCGCTGTGGGGCAGACACATAGCAGGGTGGGC	198
QY	1608	AGTGCCCTCCCTTTATCTCTGACAAATCTCTAGTCGATTTCTGCCCTTTTCTCCCGATTGGG	1667
DB	197	AGTGCCCTCCCTTTATCTCTGACAAATCTCTAGTCGATTTCTGCCCTTTTCTCCCGATTGGG	138
QY	1668	ATTTGGGGGCACGCTCTTAAGATGCGCTCTCTCCAGCCCTGTCTCAACATCTCCTCAAAATTA	1727
DB	137	ATTTGGGGGCACGCTCTTAAGATGCGCTCTCTCCAGCCCTGTCTCAACATCTCCTCAAAATTA	78
QY	1728	GTGCCAACCCAGGGGCGCTGCGACCTCCACATCATCCATTGTCTTGCTGCCAAGTCGAA	1787
DB	77	GTGCCAACCCAGGGGCGCTGCGACCTCCACATCATCCATTGTCTTGCTGCCAAGTCGAA	18
QY	1788	TAAACGGCGTGATTGCC	1804
DB	17	TAAACGGCGTGATTGCC	1
RESULT 9			
ACH14455			
ID	ACH14455	standard; cDNA; 477 BP.	
XX			
AC	ACH14455;		
XX			
DT	13-OCT-2003	(first entry)	
XX			
DE	Human adult brain cDNA #1667.		
XX			
KW	Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;		
KW	genome mapping; biodiversity; genetic disorder.		
XX			
OS	Homo sapiens.		
XX			
FN	US2003073623-A1.		
XX			
PD	17-APR-2003.		
XX			
PF	30-JUL-2001; 2001US-00918995.		
XX			
PR	30-JUL-2001; 2001US-00918995.		
XX			
PA	(DRMA/) DRMANAC R T.		
PA	(LABA/) LABAT I.		

```
PA (STAC/) STACHE-CRAIN B.  
PA (DICK/) DICKSON M C.  
XX (JONE/) JONES L W.  
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX WPI; 2003-615964/58.  
XX  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
PT mapping, in the recombinant production of protein, or in generating  
PT antisense DNA or RNA.  
XX  
XX Claim 1; SEQ ID NO 1667; 44pp; English.  
XX  
XX The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
CC determined by the technique of SBH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The present sequence  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX  
XX Sequence 477 BP; 86 A; 174 C; 125 G; 90 T; 0 U; 2 Other;  
XX  
Query Match 24.1%; Score 436.6; DB 8; Length 477;  
Best Local Similarity 99.1%; Pred. No. 7.4e-108;  
Matches 439; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 986 TGTGATGAGTCGTGTTGCTCCCATCTGAAGGCGCCAGCTTCGGGCGCCCGCGGATG 1045  
DB 35 TGGATTCGTGCTGTTGCTCCCATCTGAAGGCGCCAGCTTCGGGCGCCCGCGGATG 94  
QY 1046 GCGAAGGGGATCGCGCAAGCTCCGTGCTCTCTTGTGACGCCACCACTACCCCGAGG 1105  
DB 95 GCGAAGGGGATCGCGCAAGCTCCGTGCTCTCTTGTGACGCCACCACTACCCCGAGG 154  
QY 1106 GGTAGCCACTCGCCCGCCCGGAGGACCACTCGAGCCATTCACCCAGCTGTGCC 1165  
DB 155 GGTAGCCACTCGCCCGCCCGGAGGACCACTCGAGCCATTCACCCAGCTGTGCC 214  
QY 1166 TCCAAGACAGAGCGGGCGGCGGAGCTCCCGAAGTTATCTATGGTGGGTATACAC 1225  
DB 215 TCCAAGACAGAGCGGGCGGCGGAGCTCCCGAAGTTATCTATGGTGGGTATACAC 274  
QY 1226 TCTTCAGCCCTCGAAGCGGGACTTCCCATTCCTCACCACCTGAATGTCCCGCAG 1285  
DB 275 TCTTCAGCCCTCGAAGCGGGACTTCCCATTCCTCACCACCTGAATGTCCCGCAG 334  
QY 1286 ACTGGTTCATCAGCCAGTCCCGCCCGCCACACAAATGGGCTCAGGATCTCAGGCTTCC 1345  
DB 335 ACTGGTTCATCAGCCAGTCCCGCCCGCCACACAAATGGGCTCAGGATCTCAGGCTTCC 394  
QY 1346 ACCTCAGGGTGACCTTCCGAGGCGCCCTGGTGACTTCCAGGGCTCGCTCAGTTAGCAAT 1405  
DB 395 ACCTCAGGGTGACCTTCCGAGGCGCCCTGGTGACTTCCAGGGCTCGCTCAGTTAGCAAT 454  
QY 1406 TCAGTGCCATCTACCCCGACGAC 1428  
DB 455 TCAGTGCCATCTACCCCGACGAC 477  
RESULT 10
```

```
ACH43954  
ID ACH43954 standard; cDNA; 484 BP.  
XX  
XX ACH43954;  
AC  
DT 13-OCT-2003 (first entry)  
XX  
XX Human foetal brain cDNA #4679.  
DE  
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
KW genome mapping; biodiversity; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX  
XX US2003073623-A1.  
PN  
XX 17-APR-2003.  
PD  
XX 30-JUL-2001; 2001US-00918995.  
PF  
XX 30-JUL-2001; 2001US-00918995.  
PR  
XX (DRNA/) DRMANAC R T.  
PA (LABA/) LABAT I.  
PA (STAC/) STACHE-CRAIN B.  
PA (DICK/) DICKSON M C.  
PA (JONE/) JONES L W.  
XX  
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX  
XX WPI; 2003-615964/58.  
XX  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
PT mapping, in the recombinant production of protein, or in generating  
PT antisense DNA or RNA.  
XX  
XX Claim 1; SEQ ID NO 31166; 44pp; English.  
XX  
XX The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
CC determined by the technique of SBH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The present sequence  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX  
XX Sequence 484 BP; 118 A; 129 C; 137 G; 85 T; 0 U; 15 Other;  
XX  
Query Match 22.7%; Score 410.6; DB 8; Length 484;  
Best Local Similarity 93.9%; Pred. No. 8.2e-101;  
Matches 419; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 426 CCTGAGGATCCCGATGCTAGGACCGAGATGCTCAGCTTTATAGTGTGACCTACAT 485  
DB 39 CNGNAAANNNNNNNAGNANACNAGCCAGATGCTCAGCTTTATAGTGTGACCTACAT 98  
QY 486 GTGACTTTCACCTCAGTTTGTGATCCGTAATAATGGCAAAATTCGAAGCTACTTTCACAGTG 545  
DB 99 GTGACTTTCACCTCAGTTTGTGATCCCAAAATGGCAAGGCGAAGCTACTTTCACAGAG 158  
QY 546 CTGTTGAGGAGTAATGAATAAACAATGCTTTGTAAGCTCTTTTCAGAGGAGCCCTCGGA 605
```

Db 159 CTGTTGAGAGATTAAATGAAACAATGCTTTGTAAGCTCTTTCCAGGAGGAGCCCTCGGA 218  
Qy 606 AGCAGGCGCTGGCGGAGAGACACACTGCTGTACACAGGACACAGGAGGAGCATGAAGA 665  
Db 219 AGCAGGCGCTGGCGGAGAGACACACTGCTGTACACAGGAGGAGCATGAAGA 278  
Qy 666 CCCCCGTGAGCTGGCGGAGTGGGATGAGACCTTCGCGCTTCAGCAGCGGTCGCCGAG 725  
Db 279 CCCCCGTGAGCTGGCGGAGTGGGATGAGACCTTCGCGCTTCAGCAGCGGTCGCCGAG 338  
Qy 726 GTGCTACCGGGTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTGGGAGGCCCG 785  
Db 339 GTGCTACCGGGTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTGGGAGGCCCG 398  
Qy 786 CAGCACCCCGGCTACCCACCGGACTTCGATCCGCGCTGGTGGAGAGGCTTAACAGAA 845  
Db 399 CAGCACCCCGGCTACCCACCGGACTTCGATCCGCGCTGGTGGAGAGGCTTAACAGAA 458  
Qy 846 CCAGAGCGGTGGCGAGGAGGATCG 871  
Db 459 CCAGAGCGGTGGCGAGGAGGATCG 484

RESULT 11  
AAH12103/c  
ID AAH12103 standard; cDNA; 578 BP.  
XX  
AC AAH12103;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA clone (3'-primer) SEQ ID NO:8938.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
OS Homo sapiens.  
XX  
PN EPI074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
XX Claim 3; SEQ ID NO 8938; 2537pp + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX  
SQ Sequence 578 BP; 134 A; 142 C; 168 G; 127 T; 0 U; 7 Other;  
Query Match 18.9%; Score 341.8; DB 4; Length 578;  
Best Local Similarity 93.6%; Pred. No. 3.8e-82;  
Matches 396; Conservative 0; Mismatches 23; Indels 4; Gaps 4;  
Qy 1392 GCTCAGTTAGCATTTTCAGTGCCTATCCACCACAGAGGTGGGGCCACCAGAAACCAA 1451  
Db 576 GCTCAGTAAGCATTTTCAGNGCCATTTACCCCNCGACAAGGTGGGNCNCCCCAGAACCAA 517  
Qy 1452 AGCCCCCTTGAATGATACTCTTTTCATCA-CGGTTGCTATGGGGCCACGGCGACAGGT 1510  
Db 516 AGCCCCCTTGAATGATGACACTCTTTTCATCAGGGGTTCCTATGGGGCCNCGCGACAGGG 457  
Qy 1511 ATGGCCCCCTTGGC-AGGCTAGGAGGA-CATTATCATCCACAGGAAACCCAGGTATTAAAG 1568  
Db 456 TAGGCCCTTGGCAAGGGTAGGAGACCATTCATCACCAGGAAACCCAGGTATTAAAG 397  
Qy 1569 AAGCCCCCTGGGGCAGACAGACATAGCAGGGGTGGGAGTTCCTCCCTTTATCCTGAC 1628  
Db 396 AAGCCCCCTGGGGCAGACAGACATAGCAGGGGTGGGAGTTCCTCCCTTTATCCTGAC 338  
Qy 1629 AATCTCTAGTCGATTCTTGCCCTTTTCTCCCGATTGCGGATTGGGGGCCACCTCTAAGA 1688  
Db 337 AATCTCTAGTCGATTCTTGCCCTTTTCTCCCGATTGCGGATTGGGGGCCACCTCTAAGA 278  
Qy 1689 TGCCTCTCTCAGCCCTCTCTCAACCATCTCTCAAAATTAGTCCCAACCCAGGGGCTGGC 1748  
Db 277 TGCCTCTCTCAGCCCTCTCTCAACCATCTCTCAAAATTAGTCCCAACCCAGGGGCTGGC 218  
Qy 1749 ACCTCCACATCATCATCTGCTTCTGCTCCCAAGTCCGAATAAAGCGCTGATTGCCAAC 1808  
Db 217 ACCTCCACATCATCATCTGCTTCTGCTCCCAAGTCCGAATAAAGCGCTGATTGCCAAC 158  
Qy 1809 TGG 1811  
Db 157 TGG 155

RESULT 12  
AAS75218  
ID AAS75218 standard; cDNA; 440 BP.  
XX  
AC AAS75218;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #11022.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX



```
Query Match          7.6%; Score 138; DB 9; Length 494;
Best Local Similarity 72.8%; Pred. No. 5.6e-27;
Matches 177; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 933 TCACCAAGGAGAGAGAAACAATACAGACCCATCAGCCACACCCCGTCTTACTGTGATG 992
DB 494 TCACNCCAGGAGAGAGAAACAATACAGGTTGATTGGCCACACCCCATCTACTGTGATG 435
QY 993 AGTCGCTGTTGGCTCCCGATCTGAAGGCGCCAGCTTCGGGGCCCGCGGATGGGGAAGG 1052
DB 434 AGTCACCTGTTGGCTCCCGCCACAGGCGACCCAGAGAGGTCTCGGACAGCCGCTGG 375
QY 1053 GGGATGCGCGAAGCTCCGCTCTCTTGTGGAGCCACACCTACCCCGAGGGGTAGCC 1112
DB 374 AGGATGGGCGAAGCTCCGACACCTTTCTGGACCCACACGAGCACCCCTAGGGGAGCC 315
QY 1113 ACTCGCCCCCGCCAGGAGGACCACTGGAGCCATTACCCAGCTGTCTCCCTCAAGA 1172
DB 314 ACTACCTCGCCCCAGGAGACCCCACTTCGGGCCATATTTTCCTAGTCCCCACAGCT 255
QY 1173 CAG 1175
DB 254 CAG 252

RESULT 14
ID ABN79266
XX
AC ABN79266;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human ORF4213 cDNA, SEQ ID NO:8425.
XX
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibitor; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW vasotrophic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiast; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
OS Homo sapiens.
XX
EN W0200190366-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-US017076.
XX
PR 24-MAY-2000; 2000US-0206690P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Leach MD, Shimkets RA;
XX
DR WPI; 2002-106200/14.
DR P-FSDS; ABP35240.
XX
XX
Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation.
XX
XX
Claim 1; Page 2350; 2508pp; English.
PS
XX
```

```
Sequences ABP31028-ABP35561 represent 4534 novel human proteins
designated ORF (open reading frame) 1-4534, and sequences ABN75054-
ABN79587 represent cDNAs encoding them. The invention also encompasses
polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
referred to as ORFX) proteins, polynucleotides at least 85% identical to
the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
polynucleotides, the recombinant production of ORFX proteins, antibodies
specific for ORFX proteins, methods of detecting ORFX polynucleotides and
polypeptides, methods of screening for modulators of ORFX expression or
activity, and methods of screening individuals for a predisposition to an
ORFX-associated disorder. The ORFX proteins of the invention have a wide
range of biological activities, such as cytokine, cell proliferation,
cell differentiation, immune modulation, haematopoiesis regulation,
tissue growth, angiogenesis, activin or inhibitor activity, chemotactic/
chemokinetic activity, haemostatic activity, thrombolytic activity,
receptor/ligand, antiinflammatory activity, tumour inhibition activity,
and antiinfective activity, and may also be involved in the determination
of bodily characteristics, fertility and behaviour. ORFX proteins,
nucleic acids and antibodies may be used in the treatment of cancers,
other proliferative disorders such as psoriasis and benign tumours,
neurological disorders such as epilepsy and Alzheimer's disease,
cardiovascular diseases, immune system disorders, disorders related to
organ transplantation, disorders of tissue growth and regeneration,
diseases such as diabetes mellitus, hypothyroidism, and cholesterol
storage disease, and infectious diseases caused by viral, bacterial,
fungal and other pathogens. ORFX nucleic acids may also be used as a
source of primers and probes, in the detection of ORFX genomic sequences
or transcripts, in the identification and cloning of homologous
sequences, in genetic diagnosis, and in forensic biology. The ORFX
nucleic acids may additionally be used to produce transgenic animals
which may be useful for studying the function and/or activity of ORFX
protein, and in drug screening. The ORFX proteins may also be used as
immunogens to generate specific antibodies, which are useful in the
diagnosis, treatment and monitoring of ORFX-associated diseases
XX
SQ Sequence 292 BP; 72 A; 96 C; 83 G; 41 T; 0 U; 0 Other;

Query Match          7.2%; Score 130.8; DB 6; Length 292;
Best Local Similarity 72.8%; Pred. No. 4e-25;
Matches 182; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

QY 639 CACCAAGGACACACAGGCGAGCATGAAGACCCCGCTGG-AGCTGGCGCTCAGTGGGATGCAG 697
DB 43 CACCAACACCCACTGGGCAACGTGAAGACCCCGCTGGCGCCGACGAGCGGATGAA 102
QY 698 ACCCTTGGCGCTTACGACACCGCTGGCGAGGTGGCTTACCGGGTCAAGCGCCAGACGTCATAT 757
DB 103 GCATCCACCTTCAGCACCGGACGTCGCCACAGCTACAGGGTCAAGCGCCAGGCGCATCTAT 162
QY 758 GTGGATGAGACTCTGTTTGGCAGCCCGACGACCGCCCTACCCACCGGACTTCGAT 817
DB 163 GTGGATGAGACCTCTTTGGCAGCCCGACGACCAAGAACCACTGCCCCACGACTTTGAC 222
QY 818 CCGCCCTGGGTGGAGAGGCTAACAGAACCCAGAGCGTGGGCAAGGAGGCATCGAAGGCC 877
DB 223 CCACCCCTGGGTGCAGAAATTGTAACCGATCCAGAGGAGTGGGCGCCAGGACCACCGAAGGC 282
QY 878 TTGGGGGCAA 887
DB 283 TCTCTGGCCA 292

RESULT 15
ID ABN32144
XX
AC ABN32144;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:4892.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 06:05:38 ; Search time 304 Seconds  
(without alignments)  
3305.974 Million cell updates/sec

Title: US-10-031-589-3  
Perfect score: 1811  
Sequence: 1 atgctcgccggccctcgccg.....cggctgattgccaaactgg 1811

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptcddata/2/ina/5A COMB.seq\*  
2: /cgn2\_6/ptcddata/2/ina/5B COMB.seq\*  
3: /cgn2\_6/ptcddata/2/ina/6A COMB.seq\*  
4: /cgn2\_6/ptcddata/2/ina/6B COMB.seq\*  
5: /cgn2\_6/ptcddata/2/ina/PCTUS COMB.seq\*  
6: /cgn2\_6/ptcddata/2/ina/backfiles1.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	49.8	2.7	17493	4	US-09-804-471A-3
C 2	49.8	2.7	17493	4	US-10-238-709-3
C 3	46.2	2.6	30001	1	US-08-125-468-1
C 4	46.2	2.6	30001	2	US-08-474-933-1
C 5	45.6	2.5	72928	3	US-09-009-913-1
C 6	43	2.4	4220	1	US-08-832-883-66
C 7	43	2.4	4220	2	US-08-832-877-66
C 8	43	2.4	13104	3	US-08-256-799-4
C 9	43	2.4	13104	3	US-08-462-437-4
C 10	42.4	2.3	505	4	US-09-621-976-15639
C 11	41.4	2.3	364	4	US-09-621-976-17202
C 12	41	2.3	202001	4	US-09-734-674-3
C 13	40.8	2.3	1335	5	PCT-US91-06532-1
C 14	40.6	2.2	7218	1	US-08-232-463-14
C 15	40.4	2.2	2831	2	US-08-306-713-1
C 16	40	2.2	933	4	US-09-084-669-1
C 17	39.8	2.2	6436	4	US-09-600-099-1
C 18	39.6	2.2	2721	6	US-09-526-193A-16
C 19	39.6	2.2	4768	4	US-08-232-463-14
C 20	39.6	2.2	7218	1	US-08-804-227C-13
C 21	39.6	2.2	13987	2	US-08-804-227C-13
C 22	39.6	2.2	43280	2	US-08-804-227C-13
C 23	39.6	2.2	44377	2	US-08-804-227C-7
C 24	39.6	2.2	44377	2	US-08-804-198-1
C 25	39.2	2.2	45546	4	US-09-146-053-6
C 26	39	2.2	527	4	US-09-621-976-2475
C 27	39	2.2	3524	4	US-09-077-940A-3

```

Sequence 3, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 21, Appli
Sequence 24, Appli
Sequence 269, Appli
Sequence 5, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 11, Appli

28 39 2.2 40090 4 US-09-820-004-3
c 29 39 2.2 4403785 3 US-09-103-840A-2
30 38.6 2.1 162450 4 US-09-345-882-1
c 31 38.2 2.1 8438 1 US-07-945-283-1
c 32 38.2 2.1 20284 4 US-09-526-193A-21
c 33 38 2.1 946 4 US-09-904-615-24
c 34 38 2.1 1001 4 US-09-641-638-269
c 35 38 2.1 1881 4 US-09-434-288-5
c 36 38 2.1 48763 4 US-09-916-204-3
c 37 38 2.1 80246 3 US-09-078-294-4
c 38 38 2.1 80595 3 US-09-078-294-3
c 39 38 2.1 174493 4 US-09-804-471A-3
c 40 38 2.1 174493 4 US-10-238-709-3
c 41 37.8 2.1 2263 4 US-09-595-549-1
c 42 37.8 2.1 4257 2 US-08-630-473-1
c 43 37.8 2.1 4257 3 US-09-259-821A-1
c 44 37.8 2.1 4257 3 US-08-843-659-1
c 45 37.8 2.1 12001 1 US-08-458-568A-11

ALIGNMENTS

RESULT 1
US-09-804-471A-3/c
; Sequence 3, Application US/09804471A
; Patent No. 6479269
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001164
; CURRENT APPLICATION NUMBER: US/09/804,471A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 174493
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(174493)
; OTHER INFORMATION: n = A,T,C or G
US-09-804-471A-3

Query Match 2.7%; Score 49.8; DB 4; Length 174493;
Best Local Similarity 75.8%; Pred. No. 0.0067;
Matches 75; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

QY 494 ACCTCAGTTTGTGATCGTAAATGG--ACAAATTGGAAGCTACTTCCAGTCTGTGTG 551
Db 91060 ACCTCAGTTTCTCTCAATTGTAATGGCCATAATTACCACCTACTTCAAAGAGTCTGTG 91001

QY 552 AGAGATTAAATGAACAACTCTGTAAGCTTTTGA 590
Db 91000 GAAGGATTAAATGAATAATGATGATGCTCTCAGCA 90962

RESULT 2
US-10-238-709-3/c
; Sequence 3, Application US/10238709
; Patent No. 6680188
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001164DIV
; CURRENT APPLICATION NUMBER: US/10/238,709
; CURRENT FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 4

```





```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match          2.6%; Score 46.2; DB 2; Length 30001;
Best Local Similarity 45.9%; Pred. No. 0.029;
Matches 196; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

QY 13 CTTCCGCGCGCTAGTACACGACCTGAGTGGACACGAGAGACCTCTCCATGTT 72
Db 18498 CCGCGACAGCCGGTCTGAAGCCCTCTGGAGCGCGGAGTTCGCGCTCACGTGCT 18439

QY 73 TAGGGACCTCTCGGGCTTCAGAGAGTGGCGCCCGCCCTGGCGGACTCCCGCCATCG 132
Db 18438 GCGCGCGAGAGGGAGTTCACGTCCCGCGCTTCGCGCGCCCGTGGACGACCGCTTCC 18379

QY 133 CGGCGCG---CGAATGCTCCGGGTCCGCTCCGAGTCTCTGCTGCTCTCTGTTGCTG 189
Db 18378 GGGACGTCCCGGTGGAGCGCGGACGCGGTGGGGTCCCGTCACTGTGCGCGGAGCT 18319

QY 190 GGTGCAAGTCTGGGTTCTGGATTTCGGGGCGCGTTCACAGGTAGCTGTGC 249
Db 18318 CGTCCGCGACGCGAGTGGCGGCTCGCGCGCGCTCCACAGCGGCGGACACACCATC 18259

QY 250 CGGCTCTCGGGTGAAGTCTCGCGCGCGGTGCGCGCGCGGCGGCTAGGCTCGCGGGG 309
Db 18258 GTGATCGCGCGGTGTGGCGCGCGGCGCGCGCGCGCGGAGGTGCCAGCGCTGATGT 18199

QY 310 TCCGGGCGCCAGGATTCGGGCTGACAGTTCAGCGGGATTCGGGATCACCGCGCGCC 369
Db 18198 ACTGGGCGCGCTCTACGCGCGGTGCGCGGTGAGGAGATCCGCGACGCGCGCCTGA 18139

QY 370 CCGCGCCCTCACCGACGGTCCAGACCTGTTGGGAGAGTGGGGGACGGTCCCTG 429
Db 18138 CCTGGCGCGGAGGCTGAGCCCGCGCGCTGAGCCCGCGCGTCCGCGGAGCGCGCGG 18079

QY 430 AGGATCC 436
Db 18078 AGCAGCC 18072

RESULT 5
US-09-009-913-1
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: AxyS Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 72928 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match          2.5%; Score 45.6; DB 3; Length 72928;
Best Local Similarity 62.1%; Pred. No. 0.064;
Matches 72; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 471 GTGTGACCTACATGTGACTTACCTCAGTTTTTGTGATCCGTAATAATGACAAATTCGA 530
Db 66880 GAGTGAGTTATATAAGCTTAGCATCTAAGTTTTTCTCATCTGGAAAATGGAGTTAATAC 66939

QY 531 AGCTACTTCACAGTCTGTTGAGAGGATTAATAAGAAATGCTTTAAAGCTCTT 586
Db 66940 ATCTACTGCATTTGGCTGTTGTAAGATTAAATTAACAAGATGTAAGACCT 66995

RESULT 6
US-08-832-883-66
; Sequence 66, Application US/08832883
; Patent No. 5807881
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; APPLICANT: Baldi, Alphonso
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
; OF CANCER
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,883
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8363
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-66

Query Match          2.4%; Score 43; DB 1; Length 4220;
Best Local Similarity 73.3%; Pred. No. 0.09;
```

	Matches	55;	Conservative	0;	Mismatches	20;	Indels	0;	Gaps	0;
Qy	497	TCAGTTTGTGATCCGTA	AAATGCAGAAATTCG	AAGCTACTTCACAGT	CGTGTGTGAGAGG	556				
Db	1130	TCAGTATGGTGATCTCT	AAACTGAGATATTTG	TGTTTACCTCACAG	AGCTGTTCTGAAG	1189				
Qy	557	ATTAATGAAACAAAT				571				
Db	1190	ATTAATTAAGGCAAT				1204				

```

RESULT 7
US-08-832-877-66
; Sequence 66, Application US/08832877
; Patent No. 5840506
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 116
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,877
; FILING DATE:
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-877-66

```

	Query Match	2.4%	Score 43;	DB 2;	Length 4220;
	Best Local Similarity	73.3%;	Pred. No. 0.09;		
	Matches 55;	Conservative 0;	Mismatches 20;	Indels 0;	Gaps 0;
QY	497	TCAGTTTGTGATCCGTAATAATGCACAAATTCGAAGTACTTCACAGTGTCTGTGTGAGAGG	556		
Db	1130	TCAGTATGGTGAICTCTCTAACTGGAGATATTTGTGTTTACCTCACAGAGCTGTTCTGAAG			
QY	557	ATTAAATGAAACAAT	571		
Db	1190	ATTAAATTAAGGCAAT	1204		

RESULT 8  
US-08-2366-799-4/c  
; Sequence 4, Application US/08256799  
; Patent No. 6232034  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROENQVIST, Mats  
; APPLICANT: BERGSTROEM, Sven

APPLICANT: HERNELL, Olle  
 APPLICANT: Toernell, Jan  
 TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR  
 TITLE OF INVENTION: OBTAINING THE PROTEIN AND USE THEREOF  
 NUMBER OF SEQUENCES: 26  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BROWDY AND NEIMARK  
 STREET: 419 Seventh Street, N.W., Suite 300  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 US REGISTRATION NUMBER: US/08/256,799  
 FILING DATE: 06-DEC-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: DK 88/92  
 FILING DATE: 23-JAN-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: COOPER, Iver P.  
 REGISTRATION NUMBER: 28,005  
 REFERENCE/DOCKET NUMBER: HANSSON-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-628-5197  
 TELEFAX: 202-737-3528  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 13104 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: ps459  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 1..8934  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 8866..10014  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 10511..12277  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 8835..8867  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 10015..10510  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 12278..12443  
 PS-08-256-799-4

Query Match 2.4%; Score 43; DB 3; Length 13104;  
Best Local Similarity 62.6%; Pred. No. 0.15;  
Matches 67: Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy	484	ATGTGACTTCACTCTGACTTTTGTGATCCGTAATAATCGAACAATTCGAGCTACTTCCACAG	543
Db	4542	ATCTTTTCAGCACCTCGATTTCATCCCTCTGTAGAATGAGGATACTAAATCTCTCCCTTATGG	4483
Qy	544	TGCTGTTTGACAGGATTAATATGAACAATGCTTCTAAAGCTCTTTTGA	590
Db	4482	GAGTGTTTGAAGGATCGAAGTGAGAAAATCACTGTTAACTTCTTAGCA	4436

RESULT 9  
US-08-462-437-4/c  
; Sequence 4, Application US/08462437  
; Patent No. 6232094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMQVIST, Mats  
; APPLICANT: BERGSTROM, Sven  
; APPLICANT: HERNELL, Olie  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS  
; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,437  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON-1A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13104 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: ps459  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 1..8834  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 8868..10014  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 10511..12277  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 8835..8867  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 10015..10510  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 12278..12443  
; US-08-462-437-4  
Query Match 2.4%; Score 43; DB 3; Length 13104;

Best Local Similarity 62.6%; Pred. No. 0.15;  
Matches 67; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
Qy 484 ATGTGACTTCACCTCAGTTTGTGATCCGTAATAATCGAAGCTACTTCACAG 543  
Db 4542 ATCTTCAGCACCTCAGTTTCATCTCTGTAGATGAGGATATAATCTCCCTTATGG 4483  
Qy 544 TGCTGTTGAGAGGATTAAATGAAACAATGCTTCTAAAGCTCTTTGCA 590  
Db 4482 GAGTGTGAAAGGATGAAGTGAGAAATCCATGTAACTTCTTTAGCA 4436  
RESULT 10  
US-09-621-976-15639/c  
; Sequence 15639, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976.  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 15639  
; LENGTH: 505  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-15639

Query Match 2.3%; Score 42.4; DB 4; Length 505;  
Best Local Similarity 10.8%; Pred. No. 0.05;  
Matches 37; Conservative 158; Mismatches 149; Indels 0; Gaps 0;  
Qy 1017 AAGCGCCAGCTTCGGGGCCCGCGGATGCGGAGGGGGATGCGCAAGCTCCGTGCTC 1076  
Db 370 AAARRRMAAGGSGYCGTSTSGKMTGRKSGMTKRMWYSGMWTYSYKCTTKTKY 311  
Qy 1077 TCTTGTGAGCCACCACTACCCAGGGGTAGCCACTCGCCGCCGCCAGGAGGCAC 1136  
Db 310 TGWKSXKTRWTCTSWRYMMWMSGWARSMKSWARSWYMWACWCMMSASAYRARRSMYG 251  
Qy 1137 CACTGCGAGCCATTCAACCCAGCTGTCCTCCCAAGACAGAGCGGGGCCGCGCAGACT 1196  
Db 250 ARRSMMRAGAGWWRARRGKKRARRKSSMMRSKMSMMRSAGKARWCRWMSCRMVSC 191  
Qy 1197 CCCAGAGTTATCTATGGTGGTTACACTCTTCACGCCCTCCAGCGGGGACTTTCCC 1256  
Db 190 MGSKMSCRGTCAKMWRYARYAKYASSMGKYMGRWCYAKCARMYGYRSTRSGSR 131  
Qy 1257 ATTCCTTCACCCACTGAATGTCCTCCAGACTGTGTCATCCAGCCACAGTGCCCCCACA 1316  
Db 130 GMKYRRRRYYMMRYMMWMSWMCYRMGAAYGMSARAYRYMYSMSACKMCRMMFMQMSWM 71  
Qy 1317 CAAATGGGCTCAGGATCTCAGGCTTCACAGTCCAGGGGTGACC 1360  
Db 70 WWRCSRYSRYCWMGSKWCYSCCGYCCSACRMCYCWTRMRKSWYS 27

RESULT 11  
US-09-621-976-17202/c  
; Sequence 17202, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21





This Page Blank (uspto)



GenCore version 5.1.5  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 05:02:45 ; Search time 4878 Seconds

(without alignments)  
11066.595 Million cell updates/sec

Title: US-10-031-589-3

Perfect score: 1811

Sequence: 1 agtgcctggccctcgccg.....cggcgtattgccaacctgg 1811

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gss1:*
29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1012.2	55.9	1152	13	BX358473
2	976.8	53.9	1201	13	BX339727
3	972	53.7	1201	13	BX339726
4	932.4	51.5	1196	9	AL523765

5	930.4	51.4	1201	13	BX403928
6	924.4	51.0	1201	9	AL560267
7	910.4	50.3	979	13	BX354462
8	898.6	49.6	933	13	BX345478
9	889.2	49.1	1030	9	AL582549
10	884.2	48.8	1047	12	BM903591
11	882.2	48.7	1201	9	AL582286
12	880.6	48.6	1201	9	AL517881
13	876.8	48.4	1201	13	BX384471
14	876.6	48.4	1201	13	BX376800
15	863.6	47.7	1201	13	BX332260
16	860.2	47.5	1017	12	BM921038
17	857.6	47.4	1048	9	AL561282
18	852	47.0	1201	9	AL560832
19	850.6	47.0	1201	13	BX394189
20	849	46.9	1201	13	BX353787
21	835.4	46.1	888	13	BX346001
22	830.2	45.8	934	13	BX352694
23	828	45.7	1201	9	AL581746
24	826.8	45.7	950	13	BQ943060
25	824	45.5	890	13	BX366971
26	817.6	45.1	1201	9	AL516514
27	812.4	44.9	1067	9	AL555180
28	809.6	44.7	879	13	BU526901
29	803.4	44.4	907	13	BQ962573
30	791.6	43.7	1042	13	BX366455
31	785.8	43.4	1071	13	BX356915
32	781.2	43.1	867	12	B1758555
33	778.4	43.0	848	13	BU540911
34	778.2	43.0	1029	13	BQ057743
35	778.2	43.0	1079	13	BQ278768
36	777.6	42.9	977	12	BM468165
37	766.6	42.3	964	13	BQ670315
38	765.2	42.3	841	12	B1858698
39	764	42.2	1020	12	BM559325
40	763.4	42.2	990	12	BM911750
41	761.2	42.0	799	12	B1761620
42	760	42.0	991	13	BX375937
43	758.4	41.9	849	12	BG281865
44	758.4	41.9	909	14	CF995108
45	753.8	41.6	1197	13	BX384470

#### ALIGNMENTS

RESULT 1

BX358473

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BX358473 1152 bp mRNA linear EST 02-MAY-2003  
BX358473 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
Clone CS0DI039YB15 5-PRIME, mRNA sequence.

EST.

GI:30347985

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 1152)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 6579.r

For more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DI039CA08QP1&cluster=6579.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ invitroGen Corporation 1600

Faraday Avenue Genoscope sequence ID : CSODI039CA08QP1.

# FEATURES

Location/Qualifiers  
1..1152  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI039YB15"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match  
Best Local Similarity 98.4%; Pred. No. 2.1e-232; Length 1152;  
Matches 1060; Conservative 4; Mismatches 8; Indels 5; Gaps 4;

196 AAGTCTGGGTTCTGGGTTTCTGGATTGCGGGCCGTTTCACAGTAGCCTGTGCCGGCTC 255  
Db |||||  
71 AAGTCTGGGTTCTGGGTTTCTGGATTGCGGGCCGTTTCACAGTAGCCTGTGCCGGCTC 130  
Qy |||||  
256 CTCGGGTGAGTCGTCGCGCGGGGTCGCCCGGACGGCTAGGCTGCGGGGGTCCGGG 315  
Db |||||  
131 CTCGGGTGAGTCGTCGCGCGGGGTCGCCCGGACGGCTAGGCTGCGGGGGTCCGGG 190  
Qy |||||  
316 GCGCCAGGCATTCGCGGCTCAGATTGACGGGATCCGGATGACCGGGCCCGCGG 375  
Db |||||  
191 GCGCCAGGCATTCGCGGCTCAGATTGAC--GGATCCGGATGACCGGGCCCGCGG 249  
Qy |||||  
376 CCTCACCGAGGGTCAGACTGTGTGGGAAGAGTGGCGGACCGGGTCCCTGAGGATC 435  
Db |||||  
250 CCTCACCGAGGGTCAGACTGTGTGGGAAGAGTGGCGGACCGGGTCCCTGAGGATC 309  
Qy |||||  
436 CCGATGCTACGACGACCAAGATGCTCAGCTTTATAGGTGTGACCTACACATGTGACTTCAC 495  
Db |||||  
310 CCGATGCTACGACGACCAAGATGCTCAGCTTTATAGGTGTGACCTACACATGTGACTTCAC 369  
Qy |||||  
496 CTCAGTTTGTGATCCGTAAATGGAACAATTCGAAGTACTTTCACAGTGTCTGTTGAGAG 555  
Db |||||  
370 CTCAGTTTGTGATCCGTAAATGGAACAATTCGAAGTACTTTCACAGTGTCTGTTGAGAG 429  
Qy |||||  
556 GATTAATGAAACAATGCTTTGTAAGCTCTTTTCAGAGGAGGCTTCGGAACGAGGGCT 615  
Db |||||  
430 GATTAATGAAACAATGCTTTGTAAGCTCTTTTCAGAGGAGGCTTCGGAACGAGGGCT 489  
Qy |||||  
616 GCGCGGACAGCACACCTGCTGTACAGGAGGACACAGGACGATGAAGACCCCGGTGA 675  
Db |||||  
490 GCGCGGACAGCACACCTGCTGTACAGGAGGACACAGGACGATGAAGACCCCGGTGA 549  
Qy |||||  
676 GCTGGCGGTGAGTGGATGAGACCTCGGCTTCAGACCGGTGCGAGGTGGCTACCG 735  
Db |||||  
550 GCTGGCGGTGAGTGGATGAGACCTCGGCTTCAGACCGGTGCGAGGTGGCTACCG 609  
Qy |||||  
736 GGTCAAGGCGCAGGACGTCATATGTGATGAGACTCTGTTTGGCAGCCAGCAGGACCGG 795  
Db |||||  
610 GGTCAAGGCGCAGGACGTCATATGTGATGAGACTCTGTTTGGCAGCCAGCAGGACCGG 669  
Qy |||||  
796 GCTTACCCACCGGACTTCGATCCGCTTGGTGGAGAAAGGCTTAACAGAACCCAGAGGGCT 855  
Db |||||  
670 GCTTACCCACCGGACTTCGATCCGCTTGGTGGAGAAAGGCTTAACAGAACCCAGAGGGCT 729  
Qy |||||  
856 GGCACAGGAGGCTTGAAGGCTTGGGCAAGGGGAGCTGTGAGACCCACCCCTCAAG 915  
Db |||||  
730 GGCACAGGAGGCTTGAAGGCTTGGGCAAGGGGAGCTGTGAGACCCACCCCTCAAG 789  
Qy |||||  
916 GGCACAGCACCACCCCTCAACCAAGGAAGAAACAATACAGACCCATCAGCCACAC 975  
Db |||||  
790 GGCACAGCACCACCCCTCAACCAAGGAAGAAACAATACAGACCCATCAGCCACAC 849  
Qy |||||  
976 CCGCTCTTACTGTGATGCTGCTGTTTGGCTCCCGATCTGAAGGCGGCAGCTTCGGGGC 1035  
Db |||||

850 CCCGTCCTTACTGTGATGAGTCGCTGTTTGGCTCCGATCTCAAGCGCCACTTTTCGGGG 909  
Qy |||||  
1036 CCCGCGATGCGGAAGGGGATGCGCAAAAGCTCGTCTCTTGTGGAGCCACCAACC 1095  
Db |||||  
910 CCCGCGATGCGGAAGGGGATG-CGCAAAAGCTCGTCTCTTGTGGAGCCACCAACC 968  
Qy |||||  
1096 TACCCCGAGGGGTAGCACTC--GCCCGCCCGAGGAGGACACCTGCCAGCCATTTCAC 1153  
Db |||||  
969 TACCCCGAGGGGTAGCACTCCGCCCGCCCGCCCAAGGAGGACACCTGCCAGCCATTTCAC 1028  
Qy |||||  
1154 CCAGTGTCTCCTCAAGACAGAGCCGGGCGCAGACTCCCAAGAGTTATCTATG 1213  
Db |||||  
1029 CCAGTGTCTCCTCAAGACAGAGCCGGGCGCAGACTCCCAAGAGTTATCTATD 1088  
Qy |||||  
1214 GGTGGTTTACACTCTTCCAGCCCGCTCAAGCGGGGACTTTCCCATTCCTCCACCCAC 1270  
Db |||||  
1089 GGTGGTTTACACTCTTCCAGCCCGCTTGAAGCGGGAC--TTTCCATTCCTCCACCCAC 1144

RESULT 2  
BX339727  
LOCUS  
DEFINITION BX339727 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CSODI076YG21 5-PRIME, mRNA sequence.  
BX339727  
ACCESSION BX339727  
VERSION BX339727.1 GI:30335832  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6579.r For more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODI076AD11QPI&cluster=6579.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSODI076AD11QPI.

FEATURES  
source  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI076YG21"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match  
Best Local Similarity 99.0%; Pred. No. 7.2e-224; Length 1201;  
Matches 1008; Conservative 5; Mismatches 2; Indels 3; Gaps 3;

762 ATGAGACTCTGTTTGGCAGCCCGAGGACCCCGGCTACCCACCGGACTTCGATCCGC 821  
Qy |||||  
69 ATGAGACTCTGTTTGGCAGCCCGAGGACCCCGGCTACCCACCGGACTTCGATCCGC 128  
Db |||||  
822 CCTGGGTGGAGAGGCTTAACAGAACCCAGAGGGCTGGGCAAGGAGGATCGAAGGCTTGG 881  
Qy |||||  
129 CCTGGGTGGAGAGGCTTAACAGAACCCAGAGGGCTGGGCAAGGAGGATCGAAGGCTTGG 188  
Db |||||

882 GGGCAAAAGGGAGCTGTGAGACCAACCCCTCAAGGGGAGCAGCCACCCCAACCTCACACCA 941  
QY  
189 GGGCAAAAGGGAGCTGTGAGACCAACCCCTCAAGGGGAGCAGCCACCCCAACCTCACACCA 248  
Db  
942 GGAAGAGAAAGAAATACAGACCCATCAGCAGACCCCTCTACTGTGATGAGTCGCTGT 1001  
QY  
249 GGAAGAGAAAGAAATACAGACCCATCAGCAGACCCCTCTACTGTGATGAGTCGCTGT 308  
Db  
1002 TTGGCTCCGATCTCTGAGGGGAGCTTCCGGGGCCCGGGATGCGAAGGGGATGCGG 1061  
QY  
309 TTGGCTCCGATCTCTGAGGGGAGCTTCCGGGGCCCGGGATGCGAAGGGGATGCGG 368  
Db  
1062 CAAAGCTCGCTCTCTTTGTGAGGCGACACCTACCCAGGAGGAGGATGAGCAGTGGCCCC 1121  
QY  
369 CAAAGCTCGCTCTCTTTGTGAGGCGACACCTACCCAGGAGGAGGATGAGCAGTGGCCCC 428  
Db  
1122 GCCCAGGAGGAGCAGCAGCTGCGAGGAGCTTACCCAGGAGTGGCTCTCCCAAGACAGAGCGG 1181  
QY  
429 GCCCAGGAGGAGCAGCAGCTGCGAGGAGCTTACCCAGGAGTGGCTCTCCCAAGACAGAGCGG 488  
Db  
1182 GGCAGGCGCAGACTCCCAAGAGTTATCTATGGGTGGGTACACTCTTCAAGCCGCCCTGA 1241  
QY  
489 GGCAGGCGCAGACTCCCAAGAGTTATCTATGGGTGGGTACACTCTTCAAGCCGCCCTGA 548  
Db  
1242 AGCGGGGACTTTCCGATTCCTTCCAGCAGCTGAATGTCGCCAGCAGCTGTCATCCAGCCA 1301  
QY  
549 AGCGGGGACTTTCCGATTCCTTCCAGCAGCTGAATGTCGCCAGCAGCTGTCATCCAGCCA 608  
Db  
1302 CAGATGCCCCCAGCAGAAATGGGCTCAGGATCTCAGGCTTCCACGTCAGGGGTGACCT 1361  
QY  
609 CAGATGCCCCCAGCAGAAATGGGCTCAGGATCTCAGGCTTCCACGTCAGGGGTGACCT 668  
Db  
1362 TCGGAGGCCCTGTGTGATTCAGGGCTCGCTAGTTAGATTTTCAGTGCCATCAGCC 1421  
QY  
669 TCGGAGGCCCTGTGTGATTCAGGGCTCGCTAGTTAGATTTTCAGTGCCATCAGCC 728  
Db  
1422 CAGCAGAGTGGGGCCAGCAGAAACCAAGCCCTTGGAAATGATCTTTTCATCA 1481  
QY  
729 CAGCAGAGTGGGGCCAGCAGAAACCAAGCCCTTGGAAATGATCTTTTCATCA 788  
Db  
1482 GGGTTGCTATGGGGCCAGCAGAGTATGGCCCTTGGCAGGGTAGGAGCAATCA 1541  
QY  
789 GGGTTGCTATGGGGCCAGCAGAGTATGGCCCTTGGCAGGGTAGGAGCAATCA 848  
Db  
1542 TCACCCAGGAGCCAGGATTAAGAGCCCTTGGGGGAGCAGCAGATAGCAGG 1601  
QY  
849 TCACCCAGGAGCCAGGATTAAGAGCCCTTGGGGGAGCAGCAGATAGCAGG 907  
Db  
1602 GTGGCAGTGCCTCTTATCTGACAAATCTAGTCCGATCTTGGCCCTTTCTCCCGA 1661  
QY  
908 GTGGCAGTGCCTCTTATCTGACAAATCTAGTCCGATCTTGGCCCTTTCTCCCGA 967  
Db  
1662 TTGCGGATTTGGGGCCAGCTCTAAGATGCTCTTCCAGCCCTGTCTCAACCATCTCC 1721  
QY  
968 TTGCGGATTTGGGGGAGCCTCTAAGATGCTCTTCCAGCCCTGTCTCAACCATCTCC 1027  
Db  
1722 AATATTAGTCCAAACCCAGGGGCTGACCTCCACATCATCATCTGCTGCTGCA 1779  
QY  
1028 AATATTAGTCCAA-CCAGGGGCTGGACCT-CCACATATCATCTGCTGCTGCTCA 1083  
Db

RESULT 3  
BX339726/c 1201 bp mRNA linear EST 02-MAY-2003  
LOCUS  
DEFINITION  
BX339726 Homo sapiens, PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CSODI076YG21 3-PRIME, mRNA sequence.  
ACCESSION  
BX339726  
VERSION  
BX339726.1 GI:30333823  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL  
COMMENT  
Contact: Genoscope  
Genoscope, Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6579.r. For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODI076AD11NPL&cluster=6579.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CSODI076AD11NPL.  
FEATURES  
Location/Qualifiers  
1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI076YG21"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
Query Match 53.7%; Score 972; DB 13; Length 1201;  
Best Local Similarity 97.2%; Pred. No. 1e-222;  
Matches 967; Conservative 17; Mismatches 11; Indels 0; Gaps 0;  
QY 789 GCACCCGGCTACCCACCGGACTTCGATCCGCCCTCGGTGGAGAGGCTTAACAGACCA 848  
Db 1002 GGMACCSGCTTWCCTCCACCGGACTTCATCCGCTTGGGTGGAGAGGCTTAACARAACMA 943  
QY 849 GAGCGTGGGCAAGGAGGATCGAAGGCTTGGGGGCAAGGGGAGCTGTGAGACCAACC 908  
Db 942 GAGCGTGGGCAAGGAGGATCGAAGGCTTGGGGGCAAGGGGAGCTGTGAGACCAACC 883  
QY 909 CCTCAAGGGGAGCAGCCCAACCCCTCACCAAGAGAGAGAAACAAATACAGACCAATCA 968  
Db 982 CCTCAAGGGGAGCAGCCCAACCCCTCACCAAGAGAGAGAAACAAATACAGACCAATCA 823  
QY 969 GCCACACCCGCTTCTACTGTGATGAGTCGCTTGTGGCTCCGATCTGAGGGGCCAGCT 1028  
Db 822 GCCACACCCGCTTCTACTGTGATGAGTCGCTTGTGGCTCCGATCTGAGGGGCCAGCT 763  
QY 1029 TCGGGGCCCCCGGATGCGAAGGGGATGCGCAAGGCTCCGATCTGAGGGGCCAGCT 1088  
Db 762 TCGGGGCCCCCGGATGCGAAGGGGATGCGCAAGGCTCCGATCTGAGGGGCCAGCT 703  
QY 1089 CACCACTTACCCCGAGGGGTAGCCACTCGCCCGCCCGAGGGAGGACCACTGCGAGCCA 1148  
Db 702 CACCACCTTACCCCGAGGGGTAGCCACTCGCCCGCCCGAGGGAGGACCACTGCGAGCCA 643  
QY 1149 TTCACCCAGCTGTGCTCTCCCAAGACAGAGCGGGGCGCAGCGGAGAGTATCCAGAGTTAT 1208  
Db 642 TTCACCCAGCTGTGCTCTCCCAAGACAGAGCGGGGCGCAGCGGAGAGTATCCAGAGTTAT 583  
QY 1209 CTATGGGTGGTTTACACTTTCACGCCCTTGAAGGGGGAGCTTCCGATTCCTCCACCC 1268  
Db 582 CTATGGGTGGTTTACACTTTCACGCCCTTGAAGGGGGAGCTTCCGATTCCTCCACCC 523  
QY 1269 ACCTGAATGTCCAGAGCTGTGTCATCAGCCAGCTGCCCCCAGCAACAATGGGCTC 1328  
Db 522 ACCTGAATGTCCAGAGCTGTGTCATCAGCCAGCTGCCCCCAGCAACAATGGGCTC 463  
QY 1329 AGGATCTCAGGCTTCCACGTCAGGGGTGACCTTCCGAGGCCCTCGTGACTTCCAGGG 1388  
Db 462 AGGATCTCAGGCTTCCACGTCAGGGGTGACCTTCCGAGGCCCTCGTGACTTCCAGGG 403

QY 1389 CTGCTCAGTTAGATTTCAGTGCCATCTACCCACGACGAGGTGGGCGACCCAGAAAC 1448  
 Db 402 CTGCTCAGTTAGATTTCAGTGCCATCTACCCACGACGAGGTGGGCGACCCAGAAAC 343  
 QY 1449 CAAAGCCCCCTTGGAAATGATCTCTTCATCAGGGTTCCTATGGGCGACGGGACAG 1508  
 Db 342 CAAAGCCCCCTTGGAAATGATCTCTTCATCAGGGTTCCTATGGGCGACGGGACAG 283  
 QY 1509 GTATGGCCCCCTTCCAGGCTAGGAGGACATTCATCACCAGGGAACCCAGGTATTAAAG 1568  
 Db 282 GTATGGCCCCCTTCCAGGCTAGGAGGACATTCATCACCAGGGAACCCAGGTATTAAAG 223  
 QY 1569 AAGCCCCCTTGGGCGACAGACATAGCAGGGTGGGCGAGTCCCTTTATCTCTGAC 1628  
 Db 222 AAGCCCCCTTGGGCGACAGACATAGCAGGGTGGGCGAGTCCCTTTATCTCTGAC 163  
 QY 1629 AATCTCTAGTCGATTCTTGCCTTTTCTCCGATTGCGGATTGGGGCGACCTCTAAGA 1688  
 Db 162 AATCTCTAGTCGATTCTTGCCTTTTCTCCGATTGCGGATTGGGGCGACCTCTAAGA 103  
 QY 1689 TGCTCTCTCCAGCCCTGTCTAACATATCTCAAAATTAGTCCCAACCCAGGGGCTGGC 1748  
 Db 102 TGCTCTCTCCAGCCCTGTCTAACATATCTCAAAATTAGTCCCAACCCAGGGGCTGGC 43  
 QY 1749 ACCTCCACATCATCTGCTGCTGCGCAAGTG 1783  
 Db 42 ACCTCCACATCATCTGCTGCTGCGCAAGTG 8

## RESULT 4

AL523765 1196 bp mRNA linear EST 22-MAY-2003  
 LOCUS AL523765 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
 DEFINITION cDNA clone CS0DC003Y203 5-PRIME, mRNA sequence.  
 ACCESSION AL523765  
 VERSION AL523765.2 GI:31042026  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1. (bases 1 to 1196)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On Feb 13, 2001 this sequence version replaced gi:12787258.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 6579.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DC003AC020P1&cluster=6579.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ InvitroGen Corporation 1600  
 Paraday Avenue Genoscope sequence ID : CS0DC003AC020P1.  
 Location/Qualifiers  
 1. .1196  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DC003Y203"  
 /tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES  
source

Query Match 51.5%; Score 932.4; DB 9; Length 1196;  
 Best Local Similarity 98.3%; Pred. No. 3.5e-213;  
 Matches 956; Conservative 6; Mismatches 8; Indels 3; Gaps 2;  
 QY 157 TCCGACGTGCTGCTGGCTGCTCCCTGCTGGTGCAGAGTGCCTGGGTTCTGGGTTTC 216  
 Db 65 TCCGACGTGCTGCTGGCTGCTCCCTGCTGGTGCAGAGTGCCTGGGTTCTGGGTTTC 124  
 QY 217 TGGATTCCGGGGCGGTTTACACGTAGCTGTGCGGGTCCCTCGGGTCAAGTCCCGCGC 276  
 Db 125 TGGATTCCGGGGCGGTTTACACGTAGCTGTGCGGGTCCCTCGGGTCAAGTCCCGCGC 184  
 QY 277 GCGGTGCCCCGGGACGGCTTAGCTGCGGGGGTCCGGGGCCCCAGGCATTCCGGGCTGC 336  
 Db 185 GCGGTGCCCCGGGACGGCTTAGCTGCGGGGGTCCGGGGCCCCAGGCATTCCGGGCTGC 244  
 QY 337 AGATTGACGGGGATCCCGGATGCACCGCGCGCCCCCGCGCCCTCACCGAGGGTCCAGAC 396  
 Db 245 AGATTGACGGGGATCCCGGATGCACCGCGCGCCCCCGCGCCCTCACCGAGGGTCCAGAC 304  
 QY 397 CTGCTGGGAAGAAGTTCGGGGACGGGTCCCTGAGGATCCCGATGCCCTAGAGCCAGAT 456  
 Db 305 CTGCTGGGAAGAAGTTCGGGGACGGGTCCCTGAGGATCCCGATGCCCTAGAGCCAGAT 364  
 QY 457 GCTCAGCTTTATAGGTGTGACCTACACATGTGACTTCACTCAGTTTGTGTATCCGTA 516  
 Db 365 GCTCAGCTTTATAGGTGTGACCTACACATGTGACTTCACTCAGTTTGTGTATCCGTA 424  
 QY 517 ATGGACAAATTCGAAGCTACTTCAAGTGTGTGTGAGAGGATTAATGAACAATGCTTG 576  
 Db 425 ATGGACAAATTCGAAGCTACTTCAAGTGTGTGTGAGAGGATTAATGAACAATGCTTG 484  
 QY 577 TAAAGCTTTTTCAGGAGGGAGCTCGGAGCAGGCGCTGGCGCGCAGAGCACACCTGCT 636  
 Db 485 TAAAGCTTTTTCAGGAGGGAGCTCGGAGCAGGCGCTGGCGCGCAGAGCACACCTGCT 544  
 QY 637 GTCACGAGGACACACAGGAGCATGAAGACCCCGTGGAGCTGGCGCTCAGTGGGATGA 696  
 Db 545 GTCACGAGGACACACAGGAGCATGAAGACCCCGTGGAGCTGGCGCTCAGTGGGATGA 604  
 QY 697 GACCTCGGCTTCAGACACCGCTGCGAGGTGCTACCGGGTCAAGGCCAGGACGTCTATA 756  
 Db 605 GACCTCGGCTTCAGACACCGCTGCGAGGTGCTACCGGGTCAAGGCCAGGACGTCTATA 664  
 QY 757 TGTGGATGACACTCTGTTTGGCAGCCCGCAGGACCCCGCTACCGCGCTCAAGCCACCTCA 816  
 Db 665 TGTGGATGACACTCTGTTTGGCAGCCCGCAGGACCCCGCTCAAGCCACCTCA 724  
 QY 817 TCCGCCCTGGTGGAGAGGCTTAACAGAACAGAGCGTGGGCAAGAGGCATCGAAGGC 876  
 Db 725 TCCGCCCTGGTGGAGAGGCTTAACAGAACAGAGCGTGGGCAAGAGGCATCGAAGGC 784  
 QY 877 CTTGGGGCAAGGGGAGCTGTGAGACACCCCTCAAGGGGACGACCCGCCACCTCAC 936  
 Db 785 CTTGGGGCAAGGGGAGCTGTGAGACACCCCTCAAGGGGACGACCCGCCACCTCAC 844  
 QY 937 ACCAAGGAAGAAGCAAAATACAGACCCCATCAGCCACACCCCGCTTACTGTGATGATC 996  
 Db 845 ACCAAGGAAGAAGCAAAATACAGACCCCATCAGCCACACCCCGCTTACTGTGATGATC 904  
 QY 997 GCTGTTTGGCTCCGATCTGAAGGCGCGAGCTTGGGGCCCCCGGATGGGCAAGGGGA 1056  
 Db 905 GCTGTTTGGCTCCGATCTGAAGGCGCGAGCTTGGGGCCCCCGGATGGGCAAGGGGA 963  
 QY 1057 TGCCGCAAAAGCTCCGTCCTCTTTGTGGACGCGCACACCTACCCCGAGGGGTAGCACTC 1116  
 Db 964 TGCCGCAAAAGCTCCGTCCTCTTTGTGGACGCGCAC--ACCTACCCCGAGGGGTAGCACTC 1021  
 QY 1117 GCCCGCCCCCAGG 1129  
 Db 1022 GCCCGCCCCAAGG 1034

## ORIGIN

```

RESULT 5
BX403928      1201 bp      mRNA      linear      EST 15-MAY-2003
LOCUS        BX403928 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
DEFINITION   CLOB006ZC02 5-PRIME, mRNA sequence.
ACCESSION    BX403928
VERSION      BX403928.1 GI:30766456
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        1 (bases 1 to 1201)
JOURNAL      Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
COMMENT      Full-length cDNA libraries and normalization
              Unpublished (2001)
              Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - France
              Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
              Library was constructed by Life Technologies, a division of
              Invitrogen. This sequence belongs to sequence cluster 6579.r For
              more information about this cluster, see
              http://www.genoscope.cns.fr/
              cgi-bin/cluster.cgi?seq=CLOB006ZC02RP1&cluster=6579.r. Contact :
              Feng Liang Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
              Paraday Avenue Genoscope sequence ID : CLOB006ZC02RP1.
FEATURES     Location/Qualifiers
             1..1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CLOB006ZC02"
                /tissue_type="NEUROBLASTOMA"
                /clone_lib="Homo sapiens NEUROBLASTOMA"
                /note="vector: pcwvSPORT_6; 1st strand cDNA was primed
                with a NotI-oligo(dT) primer. Five prime end enriched,
                double-strand cDNA was digested with Not I and cloned into
                the Not I and EcoRV sites of the pcwvSPORT 6 vector.
                Library was not normalized."
ORIGIN
Query Match      51.4%; Score 930.4; DB 13; Length 1201;
Best Local Similarity 96.3%; Pred. No. 1.1e-212;
Matches 1019; Conservative 9; Mismatches 15; Indels 15; Gaps 7;

QY 213 TTCTGATTCGCGGCGCTTACACGTAGCCTGTCGCGGCTCTCGGGTGAGTCGTC 272
DB |||||
DB 36 TTCTGATTCGCGGCGCTTACACGTAGCCTGTCGCGGCTCTCGGGTGAGTCGTC 95
QY 273 GCGCGGGTCCCGGGAGCGGCTAGGCTGCGCGGGGTCGCGGGCCCGAGGCATTCGGG 332
DB |||||
DB 96 GCGCGGGTCCCGGGAGCGGCTAGGCTGCGCGGGGTCGCGGGCCCGAGGCATTCGGG 155
QY 333 CTCGAGATTCACGGGATCCCGATGACCGCGCGCGCGCGCTTACCGACGGTCC 392
DB |||||
DB 156 CTCGAGATTCACGGGATCCCGATGACCGCGCGCGCGCGCTTACCGACGGTCC 215
QY 393 AGACCTGTCGGGAAGAGTGGCGGACGGGTCCTGAGGATCCCGATCGCTACGAGCCA 452
DB |||||
DB 216 AGACCTGTCGGGAAGAGTGGCGGACGGGTCCTGAGGATCCCGATCGCTACGAGCCA 275
QY 453 AGATGCTACGCTTTAAGTGTGACCTACACATGTGACCTACCTCAGTTTGTGATCG 512
DB |||||
DB 276 AGATGCTC-----AGGTGTGACCTACACATGTGACCTACCTCAGTTTGTGATCG 327
QY 513 TAAATGGCAAAATCGAAGCTACTTACAGTCTGTGTGAGAGATTAAATGAAACAATG 572
DB |||||
DB 328 TAAATGGCAAAATCGAAGCTACTTACAGTCTGTGTGAGAGATTAAATGAAACAATG 387
QY 573 CTTGTAAGCTCTTTCAGGAGGAGCCTCGGAAGCAGGCGCTTGGCGGAGCACACC 632
DB |||||

```

Feng Liang Email : fliang@lifetech.com URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DG007BE09QPl.

## FEATURES

source

1. .1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DG007YI18"  
 /tissue\_type="B CELLS (RAMOS CELL LINE)"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
 /notes="vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."

## ORIGIN

Query Match 51.0%; Score 924.4; DB 9; Length 1201;  
 Best Local Similarity 95.6%; Pred. No. 3e-211;  
 Matches 1000; Conservative 4; Mismatches 35; Indels 7; Gaps 5;

```

Qy 15 TCGCGCCCTAGTACACACGCTGAGTGGTGGACACGAGGACCTCTCCATGTTTA 74
Db 60 TCGCGCCCTAGTACACACGCTGAGTGGTGGACACGAGGACCTCTCCATGTTTA 119
Qy 75 GGGACCTCTCCGCGCTCAGGAGCGTGGCGCCCGCTGGCGGAGCTCCCGCCATCCGCG 134
Db 120 GGGACCTCTCCGCGCTCAGGAGCGTGGCGCCCGCTGGCGGAGCTCCCGCCATCCGCG 179
Qy 135 GGGCGGAATCGTCGGGTCCGCTCCGAGTGTCTGCTGCTCTCCCTGCTGTTGCTGGTGC 194
Db 180 GGGCGGAATCGTCGGGTCCGCTCCGAGTGTCTGCTGCTCTCCCTGCTGTTGCTGGTGC 239
Qy 195 AAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 254
Db 240 AAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299
Qy 255 CTTGCGGTGAGTCCGTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 314
Db 300 CTTGCGGTGAGTCCGTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359
Qy 315 GGGCCAGGATTCGCGGCTGAGATTCGCGGATTCGCGGATTCGCGGATTCGCGGATTCGCGG 374
Db 360 GGGCCAGGATTCGCGGCTGAGATTCGCGGATTCGCGGATTCGCGGATTCGCGGATTCGCGG 419
Qy 375 GCGCTCACCGACGCGGCTCCAGACCTGCTGGGAGAGAGTGGGAGCGGCTCCCTGAGGAT 434
Db 420 GCGCTCACCGACGCGGCTCCAGACCTGCTGGGAGAGAGTGGGAGCGGCTCCCTGAGGAT 479
Qy 435 CCGATGCTTACGAGCCAGATGCTCAGCTTTATAGTGTGACCTACATGTGACTTCA 494
Db 480 CCGATGCTTACGAGCCAGATGCTCAGCTTTATAGTGTGACCTACATGTGACTTCA 539
Qy 495 CTTGAGTTTGTGATCCGTAAATGACAAATTCGAAGTACTTTCAGTGTGTTGAGA 554
Db 540 CTTGAGTTTGTGATCCGTAAATGACAAATTCGAAGTACTTTCAGTGTGTTGAGA 599
Qy 555 GGATTAATGAACAATGCTTGAAGCTTTTGAAGGAGGAGCTCGGAAGCAGGCGCC 614
Db 600 GGATTAATGAACAATGCTTGAAGCTTTTGAAGGAGGAGCTCGGAAGCAGGCGCC 659
Qy 615 TGGCCGCGCAGACACCTGCTGCTACAGGAGCCACAGGAGCTGAGAGCCCGCTGG 674
Db 660 TGGCCGCGCAGACACCTGCTGCTACAGGAGCCACAGGAGCTGAGAGCCCGCTGG 719
Qy 675 AGTGCCCGTCACTGGGATGACACCTCGGCTTTGAGCAGCCTGCGGAGTGCTTACC 734
Db 720 AGTGCCCGTCACTGGGATGACACCTCGGCTTTGAGCAGCCTGCGGAGTGCTTACC 779
Qy 735 GGGTCAGGCCAGGAGCTATGTGGATGAGCTCTGTTGGCAGCCCGCAGGACGCC 794

```

```

Db 780 GGGTCAAGGCCAGGACGTCATATGTGTGATGAGACTCTGTTTGGCAGCCAGCAGCACCC 839
Qy 795 GGCCTACCCACCGGACTTCGATCCGCCCTGGGTGGAGAGGCTTAAACAGAACGAGGCG 854
Db 840 GGCCTACCCACCGGACTTCGATCCGCCCTGGGTGGAGAGGCTTAAACAGAACGAGGCG 899
Qy 855 TGGGCAAGGAGGACATCGAAGGCTTGGGGCAAGAGGAGCTGTGAGACACCCCTCAA 914
Db 900 TGGGCAAGGAGGACATCGAAGGCTTGGGGCAAGAGGAGCTGTGAGACCA-CCCCTCAA 958
Qy 915 GGGGAGACACCCACCCCTCACCAAGGAGAGAGACAAATACAGACCCATCAGCCACA 974
Db 959 GGGGAGACCA-CCCCACCCCTCACCAAGGAGAGACAAATACAGA---CCATCAGCCACA 1014
Qy 975 CCCGCTTCTACTGTGATGAGTCTGCTGCTCCGATCTGAGAGGCGGCGAGCTTCGGGG 1034
Db 1015 -CCGCTTCTACTGTGATGA-TCGCTGTTTGGTCCCGATCTGTGAAGCGCACTTCGGGGCC 1072
Qy 1035 CCCGCGGATGCGAAGGGGATGCC 1060
Db 1073 CSCGATGSGAGGGGATGCGCAASTC 1098

```

## RESULT 7

BX354462/c

LOCUS

DEFINITION

BX354462 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens

ACCESSION

BX354462

VERSION

BX354462.1 GI:30383876

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 979)

AUTHORS

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr. Web : www.genoscope.cns.fr

Library was constructed by life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 6579.r For

more information about this cluster, see

<http://www.genoscope.cns.fr/>

cgi-bin/cluster.cgi?seq=CS0DC020B04NP1&amp;cluster=6579.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :

<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DC020B04NP1.

Location/Qualifiers

1. .979

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DC020YC08"

/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoRV

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match

50.3%; Score 910.4; DB 13; Length 979;

Best Local Similarity 96.8%; Pred. No. 6.3e-208;

Matches 953; Conservative 14; Mismatches 12; Indels 5; Gaps 4;

Qy 813 TCGATCCGCTCGGTGGAGAGCTTAAAGAACAGAGAGCGGTGGGCAAGGAGGATCGA 872

Db 979 TCGATCCGCTCGGTGG--AGAGCTTAAAGAACAGAGAGCGGTGGGCAAGGAGGATCGA 923

QY 873 AGGCTTTGGGGCAAGGGAGGTGTGAGACACCCCTCAAGGGGAGCACCCCCACCC 932  
 Db 922 AGGCTTT-GGGGCAAGGGAGGTGTGAGACACCCCTCAAGGGGAGCACCCCCACCC 864  
 QY 933 TCACACCAAGGAAGAAACAAATACAGACCCATCAGCACACCCCGTCTTACTGTGATG 992  
 Db 863 TCACACCAAGGAAGAAACAAATACAGACCCATCAGCACACCCCGTCTTACTGTGATG 804  
 QY 993 AGTGGCTTTGGCTCCCGATCTGAAGGGCCAGCTTCGGGGCCCGCGGATGGCGAAGG 1052  
 Db 803 AGTGGCTTTGGCTCCCGATCTGAAGGGCCAGCTTCGGGGCCCGCGGATGGCGAAGG 744  
 QY 1053 GGGATCGCGCAAGCTCCGCTCTCTGTGGAGCGCACACCATCACCAGGGGTAGCC 1112  
 Db 743 GGGATCGCGCAAGCTCCGCTCTCTGTGGAGCGCACACCATCACCAGGGGTAGCC 684  
 QY 1113 ACTCGCCCCCCCCAGGAGGCCACCACTGGAGGCCATTACCCAGCTGGTTCCTTCAAGA 1172  
 Db 683 ACTCGCCCCCCCCAGGAGGCCACCACTGGAGGCCATTACCCAGCTGGTTCCTTCAAGA 624  
 QY 1173 CAGAGCGGGGCGAGCGGAGACTCCAGAAAGTTATCTATGGTGGGTACACTTTTAC 1232  
 Db 623 CAGAGCGGGGCGAGCGGAGACTCCAGAAAGTTATCTATGGTGGGTACACTTTTAC 564  
 QY 1233 GCCCCTGAAGCGGGACTTTCCATTCCCTCACCACCTGAATGTCCCGAGCACTGGTC 1292  
 Db 563 GCCCCTGAAGCGGGACTTTCCATTCCCTCACCACCTGAATGTCCCGAGCACTGGTC 504  
 QY 1293 ATCAGGCCACAGTGGCCCCCACCACAAATGGGCTCAGAGTCTCAGGCTTCCAGTCTAG 1352  
 Db 503 ATCAGGCCACAGTGGCCCCCACCACAAATGGGCTCAGAGTCTCAGGCTTCCAGTCTAG 444  
 QY 1353 GGGTGACCTTCGGAGCCCCCTGGTGACTCCAGGGCTCGCTCAGTACCATTTAGTGC 1412  
 Db 443 GGGTGACCTTCGGAGCCCCCTGGTGACTCCAGGGCTCGCTCAGTACCATTTAGTGC 384  
 QY 1413 CATCTACCCACAGCAGAGTGGGGCCACCCAGAAACCAAGACCCCTTGGAAATGATACT 1472  
 Db 383 CATCTACCCACAGCAGAGTGGGGCCACCCAGAAACCAAGACCCCTTGGAAATGATACT 324  
 QY 1473 CTTTCATCAGGTTGCTATGGGGCCACGGCGACAGGTATGGCCCTTCGCGGGTAGGA 1532  
 Db 323 CTTTCATCAGGTTGCTATGGGGCCACGGCGACAGGTATGGCCCTTCGCGGGTAGGA 264  
 QY 1533 GGCATTTATCATCCACAGGAAACCCAGAGTATTAAAGAGCCCTGTGGGGCGACAGAC 1592  
 Db 263 GGCATTTATCATCCACAGGAAACCCAGAGTATTAAAGAGCCCTGTGGGGCGACAGAC 204  
 QY 1593 ATAGAGGGTGGGAGTGCCTCCCTTTATCCTTGACAAATCTCTAGTGCATTTCTTGCTTT 1652  
 Db 203 ATAGAGGGTGGGAGTGCCTCCCTTTATCCTTGACAAATCTCTAGTGCATTTCTTGCTTT 144  
 QY 1653 TTCTCCGATTGGGATTTGGGGCCACCTCTAAGATGCTCTCTCCAGCCCTGCTCAA 1712  
 Db 143 TTCTCCGATTGGGATTTGGGGCCACCTCTAAGATGCTCTCTCCAGCCCTGCTCAA 84  
 QY 1713 CCATCTCCAAATAGTGCCAAACCCAGGGGCTTGGCAGCTTCCACATCATCATTTGCTTT 1772  
 Db 83 CCATCTCCAAATAGTGCCAAACCCAGGGGCTTGGCAGCTTCCACATCATCATTTG-NTT 25  
 QY 1773 GCTGCCAAGTGGAAATAACGGCG 1796  
 Db 24 GCTSCNMMSYSCGMGATMMCSR 1

## RESULT 8

LOCUS BX345478/c 993 bp mRNA linear EST 05-MAY-2003  
 DEFINITION BX345478 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
 cDNA clone CS0C024YF06 3-PRIME, mRNA sequence.  
 ACCESSION BX345478  
 VERSION BX345478.1 GI:30383001  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 993)  
 AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 6579.r  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0AC024DC03NP2&cluster=6579.r. Contact :  
 Feng liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Paraday Avenue Genoscope sequence ID : CS0AC024DC03NP2.

## FEATURES

Location/Qualifiers  
 1..993  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0C024YF06"  
 /tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
 /note="First strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 49.6%; Score 998.6; DB 13; Length 993;  
 Best Local Similarity 98.2%; Pred. No. 4.4e-205;  
 Matches 920; Conservative 0; Mismatches 14; Indels 3; Gaps 1;  
 QY 854 GTGGCAAGGAGGATCGAAGGCTTGGGGCAAGGGAGCTGTGAGACACCCCTCA 913  
 Db 943 GGTGGCAAGGAGATCGAAGGCTTGGGGCAAGGGAGCTGTGTGAGACACCCCTCA 887  
 QY 914 AGGGCAGCACCCCAACCTCACACCAAGGAAGAAACAAATACAGACCATCAGCCAC 973  
 Db 886 AGGGCAGCACCCCAACCTCACACCAAGGAAGAAACAAATACAGACCATCAGCCAC 827  
 QY 974 ACCCGCTTACTGTGATGAGTCCGTGTTGGCTCCGATCTGAGGGCCAGCTTCGGG 1033  
 Db 826 ACCCGCTTACTGTGATGAGTCCGTGTTGGCTCCGATCTGAGGGCCAGCTTCGGG 767  
 QY 1034 GCCCGCGGATGGGAAGGGGATGCCCAAGCTCCGTGCTCTCTTGTGACGCCACA 1093  
 Db 766 GCCCGCGGATGGGAAGGGGATGCCCAAGCTCCGTGCTCTCTTGTGACGCCACA 707  
 QY 1094 CTAACCCCAAGGGTAGCACTGCGCCCGCCAGGAGGACCACTCGAGGCCATTAC 1153  
 Db 706 CTAACCCCAAGGGTAGCACTGCGCCCGCCAGGAGGACCACTCGAGGCCATTAC 647  
 QY 1154 CCAGTGTCTCTCCACACAGAGCCGGGCCAGGGCAGACTCCAGAAGTTATCTATG 1213  
 Db 646 CCAGTGTCTCTCCACACAGAGCCGGGGCCAGGGCAGACTCCAGAAGTTATCTATG 587  
 QY 1214 GGTGGTTTACACTCTTACGCGCCCTGTAAGCGGGGACTTTCCCATTCCTCCACCCACTG 1273  
 Db 586 GGTGGTTTACACTCTTACGCGCCCTGTAAGCGGGGACTTTCCCATTCCTCCACCCACTG 527  
 QY 1274 AATGTCCCAAGCTGGTTCATCCAGCCACCACTGCCGCCCAACAATGGGCTCAGAT 1333  
 Db 526 AATGTCCCAAGCTGGTTCATCCAGCCACCACTGCCGCCCAACAATGGGCTCAGAT 467  
 QY 1334 CTCAGGCTTTCACAGTCAAGGGGTGACCTTCCGAGGCCCTCGTGGTACTTCCAGGGCTCGC 1393  
 Db 466 CTCAGGCTTTCACAGTCAAGGGGTGACCTTCCGAGGCCCTCGTGGTACTTCCAGGGCTCGC 407







```

RESULT 10
BM903591
LOCUS
DEFINITION AGENCOURT 6621549 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5721988
5', mRNA sequence.
ACCESSION BM903591
VERSION BM903591.1 GI:19352948
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1047)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs@mail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL2707 row: F column: 05
High quality sequence stop: 655.
Location/Qualifiers
1..1047
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5721988"
/lab_host="DH10B"
/clone_lib="NIH MGC 125"
/notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
of three ovaries, from females ranging in age from 38 to
49 yo. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 2.1 kb, insert size range 1-3.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 036."

FEATURES
source
Query Match 48.9%; Score 884.2; DB 12; Length 1047;
Best Local Similarity 96.1%; Pred. No. 1.3e-201;
Matches 926; Conservative 0; Mismatches 36; Indels 2; Gaps 2;

QY 386 CGGGTCCAGACCTGGTGGGAAGAGTGGGGACGGGTCCCTGAGGATCCCGATCCCTA 445
DB 6 CGGGATCAGACCTGGTGGGAGAGGTGGGGACGGGTCCCTGAGATCCCGATCCCTA 65

QY 446 CGAGCCAGATGCTCAGCTTTATAGTGTGAACCTACACATGTGACTTACCTCAGTTTG 505
DB 66 CGAGCCAGATGCTCAGCTTTATAGTGTGAACCTACACATGTGACTTACCTCAGTTTG 125

QY 506 TGATCCGTAAATGGACAAATTCGAAGCTACTTCACAGTGTCTTGGAGAGGATTAATGA 565
DB 126 TGATCCGTAAATGGACAAATTCGAAGCTACTTCACAGTGTCTTGGAGAGGATTAATGA 185

QY 566 AACAAATGCTTTAAAGCTCTTTTCAGAGGAGGACCTCGGAAGCAGGGCTGGCCGGCAGA 625
DB 186 AACAAATGCTTTAAAGCTCTTTTCAGAGGAGGACCTCGGAAGCAGGGCTGGCCGGCAGA 245

QY 626 GCACACTGTGTGTACACAGGACACACAGGAGATGAAGACCCCTGTGAGCTGGCCGTC 685
DB 246 GCACACTGTGTGTACACAGGACACACAGGAGATGAAGACCCCTGTGAGCTGGCCGTC 305

QY 686 AGTGGGATGAGACCCCTCGGCCCTTCAGACCCGCTGCCAGGTGGCTACCGGTCAGGGCC 745

```

## ORIGIN

```

RESULT 11
AL582286/c
LOCUS
DEFINITION AL582286 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
Homo sapiens cDNA clone CS0DL005Y024 3-PRIME, mRNA sequence.
ACCESSION AL582286
VERSION AL582286.2 GI:31320503
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 16, 2001 this sequence version replaced gi:12950119.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6579.r For
more information about this cluster, see
http://www.genoscope.cns.fr/

```

```

DB 306 AGTGGATGACAGCCCTCGGCCCTTCAGACCGCTGCGAGGTGCTACCGGGTCAAGGCC 365
QY 746 AGGACGTCATATGTGGATGAGACTCTGTTTGGCAGCCACGAGCACCACCGGCTTACCCCA 805
DB 366 AGGACGTCATATGTGGATGAGACTCTGTTTGGCAGCCACGAGCACCACCGGCTTACCCCA 425
QY 806 CCGGACTTCGATCCGCCCTGGTGGAGAAGGCTTAACAGAACCCAGAGCGTGGGCAAGGAG 865
DB 426 CCGGACTTCGATCCGCCCTGGTGGAGAAGGCTTAACAGAACCCAGAGCGTGGGCAAGGAG 485
QY 866 GCATCGAAGGCTTTGGGGCAAGGGAGCTGTGAGACCAACCCCTCAAGGGGCGAGCACC 925
DB 486 GCATCGAAGGCTTTGGGGCAAGGGAGCTGTGAGACCAACCCCTCAAGGGGCGAGCACC 545
QY 926 CCCACCTTCACACCAAGAGAAGAAATAATACAGACCCATACGCCACACCCCTCTTAC 985
DB 546 CCCACCTTCACACCAAGAGAAGAAATAATACAGACCCATACGCCACACCCCTCTTAA 605
QY 986 TGTGATGATGCTGCTTTGGCTCCCGATCTGAAGGCCCCAGCTTTCGGGGGCCCGCGGATG 1045
DB 606 TGTGATGATGCTGCTTTGGCTCCCGATCTGAAGGCCCCAGCTTTCGGNGGCCCGCGGATG 665
QY 1046 GCGAAGGGGGATGCCGCAAGCTCCGTCTCTTGTGGACGCCACACCTACCCCA-G 1104
DB 666 GCGAAGGGGGATGCCGCAAGCTCCGTCTCTTGTGGACGCCACACCTACCCCAAGG 725
QY 1105 GGGTAGCCACTCGCCCGCCCGCCAGGAGGACCACTCGGAGCCATTCACCCAGCTGGTCC 1164
DB 726 GGGTAGCCACTCGCCCGCCCGCCAGGAGGACCACTCGGAGCCATTCACCCAGCTGGTCC 785
QY 1165 CTCCAAGACAGAGCCGGGGCCAGCGGAGAGTATCTATGGTGGGTGTAC 1224
DB 786 CCTCAAGACAGAGCCGGGGCCAGCGGAGAGTATCTATGGTGGGTGTAC 845
QY 1225 CTCCTCAGCCCTCGTGAAGCGGGG-ACCTTCCCATCTCCCTCACCACCACTGAATGTCCCA 1283
DB 846 CTCCTCAGCCCTCGTGAAGCGGGGAACTTTCCAATTCCTCNACCACTGAATGTCCCA 905
QY 1284 GCATCTGGTCATCCAGCAGCAGTCCCGCCACACAAATGGGCTCAGATCTCAGGCTT 1343
DB 906 GCATCTGGTCATCCAGCAGCAGTCCCGCCACCAATGGGCTCAGATCTCAGGCTT 965
QY 1344 CCAC 1347
DB 966 TCCC 969

```



```
Db 184 GCGGTGCGCCGGACGGCTTAGGTCGCGGGGTTCGGGGCCCCAGGCAATCCGGGCTGC 243
Qy 337 AGATTGACGGGATCCCGGAT-GCACCGGC-GCCCCCGCGCCCTACCGAGCGGTCCAG 394
Db 244 AGATTGACGGGATCCCGGATAGCACCGGCGACCGCCCGCCCTCACCGAGCGGTCCAG 303
Qy 395 ACCTGTGGGAGCAAGTGGCGGAGCGGTCCCTGAGGATCCGATCCCTAGAGCCCAAG 454
Db 304 ACCTGTGGGAGCAAGTGGCGGAGCGGTCCCTGAGGATCCGATCCCTAGAGCCCAAG 363
Qy 455 ATGCTCAGCTTTATAGTGTGACCTACACATGTGACTTCACCTGATGTTTGTGATCGTA 514
Db 364 ATGCTCAGCTTTATAGTGTGACCTACACATGTGACTTCACCTGATGTTTGTGATCGTA 423
Qy 515 RAATGACAAATCGAAGTACTTACAGTGTGTTGAGAGATTAATGAACAATGCT 574
Db 424 RAATGACAAATCGAAGTACTTACAGTGTGTTGAGAGATTAATGAACAATGCT 483
Qy 575 TGTAAAGCTCTTTGACAGGAGCGCTCGGAAGCAGGCGCTGGCGGACAGACACCTG 634
Db 484 TGTAAAGCTCTTTGACAGGAGCGCTCGGAARCAAGCGCTGGCGGACAGACACCTG 543
Qy 635 CTGTACACGAGGACCAAGCAGCAGCATGAACCCCGTGGAGTGGCGCTCAGTGGATG 694
Db 544 CTGTACACGAGGACCAAGCAGCAGCATGAACCCCGTGGAGTGGCGCTCAGTGGATG 603
Qy 695 CAGACCTCGGCTTCAGACACCGCTCCGAGGTGCTACGGGTCAAGGCCAGGACGTC 754
Db 604 CAGACCTCGGCTTCAGACACCGCTCCGAGGTGCTACGGGTCAAGGCCAGGACGTC 663
Qy 755 TATGTGATGAGACTCTGTTGACGCCAGCAGCAGGACCCCGCTACCCCAACCGACTTC 814
Db 664 TATGTGATGAGACTCTGTTGACGCCAGCAGCAGGACCCCGCTACCCCAACCGACTTC 723
Qy 815 GATCCCGCTGGGTGAGAGGCTACAGAACAGAGGCGTGGCAAGGACATCGAG 874
Db 724 GATCCCGCTGGGTGAGAGGCTACAGAACAGAGGCGTGGCAAGGACATCGAG 783
Qy 875 GCCTTGGGGGCAAGGGGAGCTGTAGACACACCCCTCAAGGGGAGCAGCCCCACCTC 934
Db 784 GCCTTGGGGGCAAGGGGAGCTGTAGACACACCCCTCAAGGGGAGCAGCCCCACCTC 843
Qy 935 ACACCAAGGAGAGAAACAAATACAGACCATCAGCCACACCCCGTCTTACTGTGATGAG 994
Db 844 ACACCAAGGAGAGAAACAAATACAGACCATCAGCCACACCCCGTCTTACTGTGATGAG 903
Qy 995 TCGCTGTTGGCTCCCGATCTGAAGCGCGCAGCTTCGGGGCCCCCGGATGGCGAAGGG 1054
Db 904 TYGCTGTTGGCTCCCGATCTGAAGCGCGCAGCTTCGGGGAGCCCCCGGATGGCGAAGGG 963
Qy 1055 GATGCGGCAAGCTCCGTGCTCTTTGTGAGCGCCACCTACCCCAAGG 1105
Db 964 GATGCGGCAAG-CTCGGTGCTCTTTGTGAGCGCAACMCTACCCCAAGGG 1013
```

## RESULT 13

```
EX384471
LOCUS 1201 bp mRNA linear EST 08-MAY-2003
DEFINITION BX384471 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK010YL01 5-PRIME, mRNA sequence.
ACCESSION BX384471
VERSION BX384471.1 GI:30447270
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1201)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
CONTACT: Genoscope
```

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6579.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DK010CF01QP1&cluster=6579.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DK010CF01QP1.  
Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DK010YL01"  
/cell\_type="HELA CELLS COT 25-NORMALIZED"  
/cell\_line="HELA"  
/note="lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## FEATURES

source

## ORIGIN

Query Match 48.4%; Score 876.8; DB 13; Length 1201;  
Best Local Similarity 95.4%; Pred. No. 8.5e-200;  
Matches 963; Conservative 5; Mismatches 28; Indels 13; Gaps 6;

Qy 46 GTGACACAGAGACCCCTCTCCATGTTTAGGACCTCTGGGCTCAGGAGCGTGGCGCC 105  
Db 69 GTGACACAGAGACCCCTCTCCATGTTTAGGACCTCTGGGCTCAGGAGCGTGGCGCC 128  
Qy 106 CGCCCTCGGGCGGACTCCCCCATCCGCGGCGGGAATGGTCGGGTCCGGCTCCGCAATG 165  
Db 129 CGCCCTCGGGCGGACTCCCCCATCCGCGGCGGGAATGGTCGGGCGCGTCCGCAATG 188  
Qy 166 CTGTGGGTGCTCCCTGGTGTCTGGGTGCAAGTGTCTGGGTCTGGGTCTTCTGATTCGC 225  
Db 189 CTGTGGGTGCTCCCTGGTGTCTGGGTGCAAGTGTCTGGGTCTTCTGATTCGC 248  
Qy 226 GGGCGGTTCACACATAGCTGTGCGGCTCTCGGGTGAAGTCGTCGCGCGCGGTGCC 285  
Db 249 GGGCGGTTCACACATAGCTGTGCGGCTCTCGGGTGAAGTCGTCGCGCGCGGTGCC 308  
Qy 286 CGGAGCGCTTAGGCTGCGGGGGTCCGGGCGCCAGGCATTCGGGCTGCGATTCAGC 345  
Db 309 CGGAGCGCTTAGGCTGCGGGGGTCCGGGCGCCAGGCATTCGGGCTGCGATTCAGC 368  
Qy 346 GGGATCCCGGATGACACCGCGCGCGCGCTCAGGAGCGGTCCAGACCTCGTGGGA 405  
Db 369 GGGATCCCGGATGACACCGCGCGCGCGCTCAGGAGCGGTCCAGACCTCGTGGGA 428  
Qy 406 AGAAGGTCCGGGAGCGGTCCCTGAGGATCCCGATGCTACGAGCCCAAGATGCTCAGCTT 465  
Db 429 AGAAGGTCCGGGAGCGGTCCCTGAGGATCCCGATGCTACGAGCCCAAGATGCTCAGCTT 488  
Qy 466 TATAGGTGTACCTACATGTGATCTTCACTCAGTTTGTGATCCGTAATGACAA 525  
Db 489 TATAGGTGTACCTACATGTGATCTTCACTCAGTTTGTGATCCGTAATGACAA 548  
Qy 526 TTCGAAGCTACTTCACAGTGTGTTGAGAGATTAATGAACAATGCTCTTAAAGCTCT 585  
Db 549 TTCGAAGCTACTTCACAGTGTGTTGAGAGATTAATGAACAATGCTCTTAAAGCTCT 608  
Qy 586 TTGACGAGGAGGCTCGGAAGAGCGGCTTGGCGGAGAGACACCTGTGTCCAGG 645  
Db 609 TTGACGAGGAGGCTCGGAAGAGCGGCTTGGCGGAGAGACACCTGTGTCCAGG 668  
Qy 646 GACCAAGCAGCATGAAGACCCCGTGGAGCTGGCGTCACTGGGATGACAGCCCTCGG 705  
Db 669 GACCAAGCAGCATGAAGACCCCGTGGAGCTGGCGTCACTGGGATGACAGCCCTCGG 728

```

QY 706 CCTCAGACCGCTGCGAGTGGCTACCGGTCAGGGCAGAGCGTCATATGTGGATCA 765
Db 729 CCTTCAGACACCGCTGCGAGTGGCTACCGGTCAGGGCAGAGCGTCATATGTGGATGA 788
QY 766 GACTCTCTTTGGAGCGCCAGCAGGCAACCGGCTACCCACCGGATTCGATCCGCCCTG 825
Db 789 GACTCTCTTTGGAGCGCCAGCAGGCAACCGGCTACCCACCGGATTCGATCCGCCCTG 848
QY 826 GTTGGAGAGCTTAACGAACACAGAGCGTGGCAGAGGAGCATCGAAGGCTTTGGGGC 885
Db 849 GGTGGAGAGCTTAACGAACACAGAGCGTGGCAGAGGAGCATCGAAGGCTTTGGGGC 908
QY 886 AAGAGGAGAGCTGTGAGACCAACCCCTCAAGGGCAGCAGCCCGGCTCACACCAAGGAA 945
Db 909 AAGAGGAGAGCTGTGAGAC--ACCCCTCAAGGGGAGCA--CCCGCTCACACMAAGRAG 965
QY 946 GAGAGAAATACAGACCCCATCAGCACACCCCGTCTTACTGTGATGATGCTGTTGG 1005
Db 966 AAGACAAATACAG--CCATCAGCAC--CCCGTCTTACTGTGATGATGCTGTTGG 1020
QY 1006 CTCGCGATCTGAAGCGCAGCTTCGGGGCCCGCGGATGGCGAAGGGG 1054
Db 1021 CTCGCGATCTGAAGCGCAC-----TCGGGGCCCGCGGATGGCGAAGGGG 1064

RESULT 14
LOCUS BX376800 1201 bp mRNA linear EST 08-MAY-2003
DEFINITION BX376800 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
cDNA clone CS0DD005YP04 5-PRIME, mRNA sequence.
ACCESSION BX376800
VERSION BX376800.1 GI:30452532
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6579.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DD005DH02QP1&cluster=6579.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DD005DH02QP1.
FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DD005YP04"
/tissue_type="NEUROBLASTOMA COT 50-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 48.4%; Score 876.6; DB 13; Length 1201;
Best Local Similarity 97.0%; Pred. No. 9.5e-200;
Matches 9.5; Conservative 8; Mismatches 17; Indels 3; Gaps 3;
QY 179 CTGTGGTGTGGGTGCAAGTGTCTGGGTTCTGTGGATTCGGGGCGGTTCAAC 238

```

```

Db 63 CTGTGGTGTGGGTGCAAGTGTCTGGGTTCTGGATTCGGGGCCGTTCAAC 122
QY 239 GTAGCCTGTGCGGCTCCTCGGTTAGTCCGTCGCGCGCGGTGCCCGGGACGGCTAG 238
Db 123 GTAGCCTGTGCGGCTCCTCGGTTAGTCCGTCGCGCGCGGTGCCCGGGACGGCTAG 182
QY 299 GCTCCCGGGGTCCGGGGCCCGGCAATCCGGGCTGCAGATTGACGGGATCCCCGATG 358
Db 183 GCTCCCGGGGTCCGGGGCCCGGCAATCCGGGCTGCAGATTGACGGGATCCCCGATG 242
QY 359 CACCGCGCGCCCGCGCCCTCACCGGTCAGACCTGTGGGAAAGATGTCGGGG 418
Db 243 CACCGCGCGCCCGCGCCCTCACCGGTCAGACCTGTGGGAAAGATGTCGGGG 302
QY 419 ACGGTCCTGTAGGATCCGATCCGATACGAGCCAGATGCTCAGCTTTAGGTGTGACC 478
Db 303 ACGGTCCTGTAGGATCCGATCCGATACGAGCCAGATGCTCAGCTTTAGGTGTGACC 362
QY 479 TACACATGTGACTTCACTCAGTTTGTGATCCGTTAAATGACAAATTCGAAGTACTT 538
Db 363 TACACATGTGACTTCACTCAGTTTGTGATCCGTTAAATGACAAATTCGAAGTACTT 422
QY 539 CACAGTCTGTGTAGAGGATTAATGAACAATGCTGTGTAAGCTTTTCAGAGGGAG 598
Db 423 CACAGTCTGTGTAGAGGATTAATGAACAATGCTGTGTAAGCTTTTCAGAGGGAG 482
QY 599 CTTGGAAGCAGGCGCTGGCGGCGAGGACACCTGTGTACACAGGACCCAGGACG 658
Db 483 CTTGGAAGCAGGCGCTGGCGGCGAGGACACCTGTGTACACAGGACCCAGGACG 542
QY 659 ATGAAGACCCCGGTGGAGCTGGCGGTCAAGGCGAGGACCTGTGTGAGTGTGAGC 718
Db 543 ATGAAGACCCCGGTGGAGCTGGCGGTCAAGGCGAGGACCTGTGTGAGTGTGAGC 602
QY 719 TGCCGAGTGTGCTACCGGTCAGGGCAGGACCTGTGTGAGTGTGAGTGTGAGTGTG 778
Db 603 TGCCGAGTGTGCTACCGGTCAGGGCAGGACCTGTGTGAGTGTGAGTGTGAGTGTG 662
QY 779 AGCCAGCAGGACCCCGGCTACCCACCGGATTCGATCCGCGCTGGGTGGAGAGGCT 838
Db 663 AGCCAGCAGGACCCCGGCTACCCACCGGATTCGATCCGCGCTGGGTGGAGAGGCT 722
QY 839 AACGAGACAGAGCGCTGGGCAAGGAGGATCAAGGCGCTTGGGGCAAGGGAGCTGT 898
Db 723 AACGAGACAGAGCGCTGGGCAAGGAGGATCAAGGCGCTTGGGGCAAGGGAGCTGT 782
QY 899 GAGACACCCCGCTCAAGGGGCGAGCAGCCCGGCTCACACCAAGGAGGAGGAGGATAC 958
Db 783 GAGACACCCCGCTCAAGGGGCGAGCAGCCCGGCTCACACCAAGGAGGAGGAGGATAC 842
QY 959 AGACCCATAGCAGCAGCCCGCTTACTGTGATGATGCTGTGCTTGGTCCCGATCTGAA 1018
Db 843 AGACCCATAGCAGCAGCCCGCTTACTGTGATGATGCTGTGCTTGGTCCCGATCTGGA 902
QY 1019 GGGCCAGCTTCGGGGCCCGGATGGCGAAGGGGATGCGGCAAGGCTTCCGTGTCTC 1078
Db 903 AGGCGACGC--TCGCGGGCCCGCGATGGCGAAGGGGATGCGGCAAG--CTCCGTGTCTC 960
QY 1079 TTGTGGAGCCACCACTACCCCGAGGGGTAGCCACTGCGCCC 1121
Db 961 TTGTGGAGC--CAACACCTACCCCAAGGGTAGCAGTCCGCGGCC 1002

RESULT 15
LOCUS BX332260
DEFINITION BX332260 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DC010YM09 5-PRIME, mRNA sequence.
ACCESSION BX332260
VERSION BX332260.1 GI:30341128
KEYWORDS EST.
SOURCE Homo sapiens (human)

```

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1. (bases 1 to 1201)  
 Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 6579.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0C010AG05QPL&cluster=6579.r. Contact :  
 Peng Liang Email: filiang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DC010AG05QPL.  
 FEATURES  
 Location/Qualifiers  
 1. .1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DC010YM09"  
 /tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
 /notes="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 ORIGIN  
 Query Match 47.7%; Score 863.6; DB 13; Length 1201;  
 Best Local Similarity 98.3%; Pred. No. 1.3e-196;  
 Matches 891; Conservative 2; Mismatches 11; Indels 2; Gaps 2;  
 QY 507 GATCCGTAAATGACAAATTCGAAGCTACTTACAGTCTGTTGAGAGGATTAATGAA 566  
 Db 48 GGTCCGGAATTCGCGGATTCGAAGCTACTTACAGTCTGTTGAGAGGATTAATGAA 107  
 QY 567 ACAATGCTTGAAGCTCTTTGAGGAGGAGGCTCGGAAGCAGGGCTCGGCGGAGAG 626  
 Db 108 ACHATGCTTGAAGCTCTTTGAGGAGGAGGCTCGGAAGCAGGGCTCGGCGGAGAG 167  
 QY 627 CACACTGTGTCACAGGAGCACAGGAGCATGAAGACCCCGTGGAGCTGGCCGTCA 686  
 Db 168 CACACTGTGTCACAGGAGCACAGGAGCATGAAGACCCCGTGGAGCTGGCCGTCA 227  
 QY 687 GTGGATGCAGACCTCGGCTTCAGCACCGCTGCCGAGTGGCTACCGGGTCAAGGCCA 746  
 Db 228 GTGGATGCAGACCTCGGCTTCAGCACCGCTGCCGAGTGGCTACCGGGTCAAGGCCA 287  
 QY 747 GGACGTCAATATGTGGATGAGACTCTGTTGGCAGCCCGCAGGACCCCGGCTTACCCAC 806  
 Db 288 GGACGTCAATATGTGGATGAGACTCTGTTGGCAGCCCGCAGGACCCCGGCTTACCCAC 347  
 QY 807 CGGACTTCGATCGCCCTGGTGAGAGGCTTAACAGACCAGAGGGTGGGCAAGGAGG 866  
 Db 348 CGGACTTCGATCGCCCTGGTGAGAGGCTTAACAGACCAGAGGGTGGGCAAGGAGG 407  
 QY 867 CATCGAAGGCTTGGGGGCAAGGGAGCTGTGAGACCACCCCTCAAGGGGCGAGCACCC 926  
 Db 408 CATCGAAGGCTTGGGGGCAAGGGAGCTGTGAGACCACCCCTCAAGGGGCGAGCACCC 467  
 QY 927 CCACCTTCACACCAAGGAAGAAACAAATACAGACCCATCAGCCACACCCCGTCTTACT 986  
 Db 468 CCACCTTCACACCAAGGAAGAAACAAATACAGACCCATCAGCCACACCCCGTCTTACT 527  
 QY 987 GTGATGATCGCTGTTGGCTCCGATCTCGAGCGCCAGCTTCGGGGCCCGCGGATGG 1046  
 Db 528 GTGATGATCGCTGTTGGCTCCGATCTCGAGCGCCAGCTTCGGGGCCCGCGGATGG 587

QY 1047 CGAAGGGGATGCGCGCAAGCTCCGTGCTCTTTGTGGAGCGCCACCACTACCCCGAGG 1106  
 Db 588 CGAAGGGGATGCGCGCAAGCTCCGTGCTCTTTGTGGAGCGCCACCACTACCCCGAGG 647  
 QY 1107 GTAGCCACTGCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTACCCAGCTGGTCCCT 1166  
 Db 648 GTAGCCACTGCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTACCCAGCTGGTCCCT 707  
 QY 1167 CCAAGACAGAGCGCGGCGCCAGCGCAGACTCCCAAGAGTTATCTATGGTGGGTACACT 1226  
 Db 708 CCAAGACAGAGCGCGGCGCCAGCGCAGACTCCCAAGAGTTATCTATGGTGGGTACACT 767  
 QY 1227 CTTCAAGCCCTCTGAAAGCGGGGACTTTCCCATTCCTTCCCTCACCCACCTGAATGTCGCCAGCA 1286  
 Db 768 CTTCAAGCCCTCTGAAAGCGGGGACTTTCCCATTCCTTCCCTCACCCACCTGAATGTCGCCAGCA 827  
 QY 1287 CTGGTTCATCCAGCCACCACTGCGCCCGCCACACAAATGGGCTTCAGGATCTCAGGCTTTCCA 1346  
 Db 828 CTGGTTCATCCAGCCACCACTGCGCCCGCCACACAAATGGGCTTCAGGATCTCAGGCTTTCCA 887  
 QY 1347 CGTCAGGGGTGACCTTCGGAGCGCCCTTGCTGAGCTTCAGGGGTCTCAGTTAGCATTT 1406  
 Db 888 CGTCAGGGGTGACCTTCGGAGCGCCCTTGCTGAGCTTCAGGGGTCTCAGTTAGCATTT 945  
 QY 1407 CAGTGC 1412  
 Db 946 MAGTGC 951  
 Search completed: August 17, 2004, 09:42:22  
 Job time : 4890 secs

This Page Blank (uspto)

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 11:57:41 ; Search time 54 seconds  
(without alignments)  
1407.505 Million cell updates/sec

Title: US-10-031-589-4

Perfect score: 1437  
Sequence: 1 MKTPVELAVSGMOTLGLQHR.....SVFPPRRGGATQKPKPKWK 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseqp\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1437	100.0	269	4	AAB67330 Human neu
2	1437	100.0	269	4	AAB94540 Human pro
3	863	60.1	177	4	AAB94561 Human pro
4	239	16.6	467	4	ABB11341 Human typ
5	218	15.2	65	5	ABP35240 Human ORF
6	136.5	9.5	4019	4	AAE13839 Human lun
7	136.5	9.5	4019	7	ADD66733 Human lun
8	136.5	9.5	4019	7	ADDE6797 Human lun
9	136.5	9.5	4025	5	ABP69736 Human pol
10	136	9.5	2703	4	ABB63299 Drosophil
11	136	9.5	2703	5	ABG70019 Larval vi
12	134.5	9.4	19938	6	ABP76681 Streptom
13	134	9.3	502	6	ADA24266 Human WAS
14	132.5	9.2	470	4	ABG21932 Novel hum
15	131.5	9.2	705	4	AAM39328 Human pol
16	111.5	9.2	714	4	AAM41114 Human pol
17	129.5	9.0	791	5	ABP65089 Hypoxia-r
18	129.5	9.0	863	6	ABP98856 Human str
19	129.5	9.0	863	7	ADC31071 Human nov
20	129.5	9.0	863	7	ADA48296 Human MIC
21	129.5	9.0	863	7	ADA48308 Human MIC
22	128	8.9	504	2	AAAY24091 Human wis
23	128	8.9	566	6	ADA24286 Protein r
24	126.5	8.8	1199	7	ADD46005 Rat Prote
25	126.5	8.8	1199	7	ADE61171 Rat Prote

ALIGNMENTS

RESULT 1

AAB67330  
ID AAB67330 standard; protein; 269 AA.

XX AAB67330;  
AC  
XX 23-APR-2001 (first entry)  
DT  
XX Human neuron progenitor cell clone #2 protein.  
DE  
XX Neuron; progenitor cell; gene therapy.  
KW  
XX Homo sapiens.  
OS  
XX WO200107607-A2.  
FN  
XX 01-FEB-2001.  
PD  
XX 21-JUL-2000; 2000WO-JP004895.  
PF  
XX 23-JUL-1999; 99JP-00209817.  
PR  
XX 18-OCT-1999; 99US-0159528P.  
PR  
XX (HELI-) .HELIX RES INST.  
PA  
XX Ota T, Isogai T, Nishikawa T, Kawai Y;  
PI  
XX WPI; 2001-182791/18.  
DR  
XX New human polynucleotides, particularly DNAs, isolated from a cDNA  
PT library derived from progenitor cells, useful in gene therapy, as well as  
PT in producing proteins useful as diagnostic markers in drug development.  
XX  
XX Claim 1; Page 36-37; 54pp; English.  
PS  
XX The present invention relates to human proteins isolated from clones from  
CC neuron progenitor cells. The proteins and the DNA encoding them may be  
CC used in the preparation of treatments for diseases associated with the  
CC proteins  
XX  
XX Sequence 269 AA;

Query Match 100.0%; Score 1437; DB 4; Length 269;

Best Local Similarity 100.0%; Pred. No. 1.8e-114; Indels 0; Gaps 0;

Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPVELAVSGMOTLGLQHRCGGYRVARTSYVDLTFGSPAGTRTPPDFDPFWEXKA 60  
|||||

Db 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDFDPPWVEKA 60  
 QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTTLTPRKKNKYRPISHTPSYCDLSLFGSRSE 120  
 Db 61 NRTGVGKEASKALGAKGSCETTPSRGSTTLTPRKKNKYRPISHTPSYCDLSLFGSRSE 120  
 QY 121 GASFGAPRMAGDAAKLRALLWTPTTPRGSHSRPREAPLRAIHPAGPSKTEPGPAADS 180  
 Db 121 GASFGAPRMAGDAAKLRALLWTPTTPRGSHSRPREAPLRAIHPAGPSKTEPGPAADS 180  
 QY 181 QKLSMGGHLSRPLKRGSLHSLTHLNVSTGHPATSAHTNGPDLPSTSGVTFPSPLV 240  
 Db 181 QKLSMGGHLSRPLKRGSLHSLTHLNVSTGHPATSAHTNGPDLPSTSGVTFPSPLV 240  
 QY 241 TSARSVSISVPSTPRGGATQKPKPWK 269  
 Db 241 TSARSVSISVPSTPRGGATQKPKPWK 269

RESULT 2  
 AAB94540  
 ID AAB94540 standard; protein; 269 AA.  
 XX  
 AC AAB94540;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:15284.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
 XX  
 PS Claim 8; SEQ ID NO 15284; 2537pp + Sequence Listing; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification, where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to AA895893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

CC particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to AA895893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 269 AA;  
 SQ

Query Match 100.0%; Score 1437; DB 4; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-114;  
 Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDFDPPWVEKA 60  
 Db 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDFDPPWVEKA 60  
 QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTTLTPRKKNKYRPISHTPSYCDLSLFGSRSE 120  
 Db 61 NRTGVGKEASKALGAKGSCETTPSRGSTTLTPRKKNKYRPISHTPSYCDLSLFGSRSE 120  
 QY 121 GASFGAPRMAGDAAKLRALLWTPTTPRGSHSRPREAPLRAIHPAGPSKTEPGPAADS 180  
 Db 121 GASFGAPRMAGDAAKLRALLWTPTTPRGSHSRPREAPLRAIHPAGPSKTEPGPAADS 180  
 QY 181 QKLSMGGHLSRPLKRGSLHSLTHLNVSTGHPATSAHTNGPDLPSTSGVTFPSPLV 240  
 Db 181 QKLSMGGHLSRPLKRGSLHSLTHLNVSTGHPATSAHTNGPDLPSTSGVTFPSPLV 240  
 QY 241 TSARSVSISVPSTPRGGATQKPKPWK 269  
 Db 241 TSARSVSISVPSTPRGGATQKPKPWK 269

RESULT 3  
 AAB94561  
 ID AAB94561 standard; protein; 177 AA.  
 XX  
 AC AAB94561;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:15335.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.



XX PS Claim 8; SEQ ID NO 15335; 2537pp + Sequence Listing; English.

XX CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX CC Sequence 177 AA;

XX CC Query Match 60.1%; Score 863; DB 4; Length 177;

XX CC Best Local Similarity 99.4%; Pred. No. 11e-65; Indels 0; Gaps 0;

XX CC Matches 160; Conservative 1; Mismatches 0;

QY 1 MKTPELVAVSGMOTGLQHCRCGYRVKARTSVVDTEFLGSPAGTRTPDPDPWVEKA 60

DB 1 MKTPELVAVSGMOTGLQHCRCGYRVKARTSVVDTEFLGSPAGTRTPDPDPWVEKA 60

QY 61 NRTRGVGKEASKALGAKGSCETTPSRGSTPTLPRKNKYRPTSHPTSCDSLFGSRSE 120

DB 61 NRTRGVGKEASKALGAKGSCETTPSRGSTPTLPRKNKYRPTSHPTSCDSLFGSRSE 120

QY 121 GASFGAPRMAKGAALRALLWTPTTPPGSGHSPPRPREAPL 161

DB 121 GASFGAPRMAKGAALRALLWTPTTPPGSGHSPPRPREAPL 161

RESULT 4

ABH11341

XX AC ABB11341 standard; peptide; 467 AA.

XX AC ABB11341;

XX DT 11-JAN-2002 (first entry)

XX DE Human type II procollagen homologue, SEQ ID NO:1711.

XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor; hematopoiesis regulation; tissue growth; immunomodulator; activin; inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis; proliferation; metastasis; cancer; tumour; haematopoietic disorder; myeloid cell disorder; lymphoid cell disorder; asthma; arthritis; chronic inflammatory condition; proliferative retinopathy; atherosclerosis; coronary heart disease; arterial ischaemia; bone disorder; osteoporosis; vascular growth disorder; tissue regeneration; wound healing; infection; immune disorder; cell culture; drug screening; gene therapy; antiinflammatory; antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic; KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial; antifungal; vulnery; antiulcer.

XX OS Homo sapiens.

XX FN WO200157188-A2.

XX XX

PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US003800.

XX XX 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

XX XX (HYSE-) HYSEQ INC.

XX XX Tang YT, Liu C, Drmanac RT;

XX XX WPI; 2001-457740/49.

XX DR N-PSDB; ABA08585.

XX XX Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis and cancer.

XX PS Claim 20; Page 167-168; 1963pp; English.

XX CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis; cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention

XX CC Sequence 467 AA;

XX CC Query Match 16.6%; Score 239; DB 4; Length 467;

XX CC Best Local Similarity 29.4%; Pred. No. 6.7e-12;

XX CC Matches 94; Conservative 30; Mismatches 94; Indels 102; Gaps 17;

QY 5 VELAVSGMOTGLQHR--CRGYR-----VKARTSVVDTEFLGSPAGTRTPDPDPW 56

DB 176 MLCUAAQAQAPGLPRTIRGWRLTEPEAWRR--HRRFPWGORGAVRPPQGAAPPP 232

QY 57 VEKANRTRGVGKEASKALGAKGSCETTPSRGSTPTLTPRKNKYRPTSHPTSCY---DES 113

DB 233 SHQGRRT-----NTDPS--ATPRLT-----VMSRCLAPDLK 261

QY 114 LFGRSRSEGFAPRMAKGAALRALLWTPTTPPGSGHSPPRPREAPLRATHPAGESKTE 173

Db 262 APASGPRGWRGMSQSS-----GALLWTPPTPRGSHSPREAPLRAIHGAGSKSR 314  
 Qy 174 PGPAADSKLGMGLHSRPLKRG-----LSHSL-----THLNVPTGHPATS 216  
 Db 315 AGASGRFLPEVYIGWTLTPPEAGTFLIPSPXTXMSPALVIQPPVPTQMGRLISGLPRQG 374  
 Qy 217 APHNGQDLPEPSTGVTF-----RSLVTSRARSVSTSVPTPR-----257  
 Db 375 XP-SCAPWXL-PGLAQAFQCHLPHDBVGPFRNQSLGND-----TLSSGLPMGPRRWPL 429  
 Qy 258 ---CG--ATQKP-----KPPW 268  
 Db 430 ARVGCHSSPREQVLKKPLW 449  
 RESULT 5  
 ID ABP35240 standard; protein; 65 AA.  
 AC ABP35240;  
 DT 08-JUN-2002 (first entry)  
 DE Human ORF4213 protein, SEQ ID NO:8426.  
 KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnary;  
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;  
 KW neuroprotective; antithrombotic; anticoagulant; thrombolytic;  
 KW cardiatic; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.  
 OS Homo sapiens.  
 PN W0200190366-A2.  
 PD 29-NOV-2001.  
 PF 24-MAY-2001; 2001WO-US017076.  
 PR 24-MAY-2000; 2000US-0206690P.  
 PA (CURA-) CURAGEN CORP.  
 PI Leach MD, Shinkets RA;  
 DR WPI; 2002-106200/14.  
 DR N-PSDB; ABN79266.  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PS transplantation.  
 PS Claim 10; Page 2350; 2508pp; English.  
 CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or

CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases  
 XX  
 SQ Sequence 65 AA;  
 Query Match 15.2%; Score 218; DB 5; Length 65;  
 Best Local Similarity 71.4%; Pred.No.3.7e-11;  
 Matches 40; Conservative 4; Mismatches 12; Indels 0; Gaps 0;  
 Qy 12 MQTGLQHRRCGGYRVKARTSYVDETLFGSPAGTRPTPPDPDPWVEKANRTSGVG 67  
 Db 1 MKAHQLHRGFTSVRVKARSYVDETLFGSPARTRPAQDPDPWQVQNCNRSRGVG 56  
 RESULT 6  
 ID AAE13839 standard; protein; 4019 AA.  
 AC AAE13839;  
 DT 26-FEB-2002 (first entry)  
 DE Human lung tumour-specific protein SCC2-29.  
 KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 KW antisense-therapy; vaccine; immune response; lung cancer; SCC2-29.  
 OS Homo sapiens.  
 PN W0200172295-A2.  
 PD 04-OCT-2001.  
 PF 28-MAR-2001; 2001WO-US009991.  
 PR 29-MAR-2000; 2000US-00538037.  
 PR 05-JUN-2000; 2000US-00588937.  
 PR 18-AUG-2000; 2000US-00640878.  
 PR 22-SEP-2000; 2000US-0234517P.  
 PR 01-NOV-2000; 2000US-00704512.  
 PR 14-DEC-2000; 2000US-00738573.  
 PA (CORI-) CORIXA CORP.  
 PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CV;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 DR WPI; 2001-639201/73.  
 DR N-PSDB; AAD23451.  
 XX

PT New human lung-specific polynucleotides and polypeptides for the  
PT diagnosis and treatment of disease e.g. lung cancer.

XX Disclosure; Page 309-318; 378pp; English.

XX The invention relates to isolated lung tumour-specific proteins and their  
CC corresponding cDNA molecules. Lung tumour-specific proteins and their  
CC antigen-presenting cells are useful for stimulating and/or expanding T  
CC cells specific for a tumour protein, and for inhibiting the development  
CC of cancer. The invention also relates to a composition useful for  
CC stimulating an immune response, and for treating cancer. The lung tumour  
CC specific oligonucleotide is useful in gene therapy and for diagnosis,  
CC detection and treatment of lung cancer. The present sequence is human  
CC lung tumour-specific protein

XX Sequence 4019 AA;

Query Match 9.5%; Score 136.5; DB 4; Length 4019;  
Best Local Similarity 24.0%; Pred. No. 0.053;  
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDLFLGSPAGTR-----PTPDPDPWVEK-----ANTRGVGK 68  
Db 988 SRPLQWNETTANRPSVRLDSSSTTNDPYAKPDTFRVMTDQFPKSLGSRSPVSE 1047

QY 69 EASKALGAKGSC-----ETTPSRGSTTLTPRKKNK-----YRPSHTPS 108  
Db 1048 QTAKGPIAAGTSDHFTKPSRADVFQCRIPDSYARPLLTAPLDGSGFPKTPMQPPS 1107

QY 109 YCDESLFGSRSESGAFGAPRMAKGAALKRALLWTPTTPRGSH--SPRPREAPLRAIHP 166  
Db 1108 SQDP--YGSVSQ-----ASRRLSVDPIERPAL--TPRPIDNFHNQSNNDPYSQPLTTPH 1158

QY 167 -----AGPSK--TEPG-----PAADSKLSMGLHS-----SRPL 194  
Db 1159 AVNESFAHPSRAFSQPGTISRPTSDQPYSPGPTFRPVVDYSQSSGTARNTDYSQP- 1217

QY 195 KRGSLSHLTHLVNSTGHPATSGHTNGPDRLRSTSGVTFSPVTSRARSVSISVPST 254  
Db 1218 -----PGTPRTTVDYPSQQPQTRPSTQTDLFVTPVTHNQSHSDPYAHPPT 1264

QY 255 PRGGATQPKPP 267  
Db 1265 PRFGISVPYSQPP 1277

RESULT 7  
ADD66733  
ID ADD66733 standard; protein; 4019 AA.

XX AC ADD66733;

XX DT 15-JAN-2004 (first entry)

XX DE Human lung tumour-specific related protein, SEQ ID No 425.

XX expression control; cancer; T cell; tumour; immune; cytostatic; vaccine;  
KW human; lung tumour-specific.

XX OS Homo sapiens.

XX PN WO200292001-A2.

XX PD 21-NOV-2002.

XX PF 10-MAY-2002; 2002WO-US014975.

XX PR 11-MAY-2001; 2001US-00854133.

XX PA (CORI-) CORIXA CORP.

XX PI Lodes MJ, Wang T, Fan L, Algate PA, McNeill PD;

XX

DR WPI; 2003-120592/11.

XX New polynucleotide and polypeptide, useful for preparing a composition  
PT for diagnosing, treating or preventing cancer.

XX Disclosure; SEQ ID NO 425; 494pp; English.

XX The invention relates to a novel isolated polynucleotide comprising one  
CC of 32 47-6080 base pair sequences, given in the specification, or their  
CC complements or degenerate variants, at least 20 contiguous residues of a  
CC sequence in, or having at least 75 or 90 % identity with the isolated  
CC polynucleotide, or that hybridise with the polynucleotide. The invention  
CC further comprises: an isolated polypeptide; an expression vector  
CC comprising the polynucleotide operably linked to an expression control  
CC sequence; a host cell transformed or transfected with the expression  
CC vector; an isolated antibody or its antigen-binding fragment that  
CC specifically binds to the polypeptide; a method for detecting the  
CC presence of a cancer in a patient; a fusion protein comprising the  
CC polypeptide; an oligonucleotide that hybridises to the isolated  
CC polynucleotide under moderately stringent conditions; a method for  
CC stimulating and/or expanding T cells specific for a tumour protein; an  
CC isolated T cell population; a composition comprising a first component  
CC consisting of carriers and immunostimulants and a second component; a  
CC method for stimulating an immune response in a patient; a method for  
CC treating cancer in a patient; a method for determining cancer in a  
CC patient; a diagnostic kit comprising at least one oligonucleotide or  
CC antibody and a detection reagent comprising a reporter group; and a  
CC method for inhibiting the development of cancer in a patient. The  
CC compositions of the invention have cytostatic activity and can be used to  
CC create a vaccine. The isolated polynucleotide is useful for preparing a  
CC composition for diagnosing, treating or preventing cancer. This sequence  
CC represents a human lung tumour-specific protein relating to the  
CC invention.

XX Sequence 4019 AA;

Query Match 9.5%; Score 136.5; DB 7; Length 4019;  
Best Local Similarity 24.0%; Pred. No. 0.053;  
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDLFLGSPAGTR-----PTPDPDPWVEK-----ANTRGVGK 68  
Db 988 SRPLQWNETTANRPSVRLDSSSTTNDPYAKPDTFRVMTDQFPKSLGSRSPVSE 1047

QY 69 EASKALGAKGSC-----ETTPSRGSTTLTPRKKNK-----YRPSHTPS 108  
Db 1048 QTAKGPIAAGTSDHFTKPSRADVFQCRIPDSYARELLTAPLDGSGFPKTPMQPPS 1107

QY 109 YCDESLFGSRSESGAFGAPRMAKGAALKRALLWTPTTPRGSH--SPRPREAPLRAIHP 166  
Db 1108 SQDP--YGSVSQ-----ASRRLSVDPIERPAL--TPRPIDNFHNQSNNDPYSQPLTTPH 1158

QY 167 -----AGPSK--TEPG-----PAADSKLSMGLHS-----SRPL 194  
Db 1159 AVNESFAHPSRAFSQPGTISRPTSDQPYSPGPTFRPVVDYSQSSGTARNTDYSQP- 1217

QY 195 KRGSLSHLTHLVNSTGHPATSGHTNGPDRLRSTSGVTFSPVTSRARSVSISVPST 254  
Db 1218 -----PGTPRTTVDYPSQQPQTRPSTQTDLFVTPVTHNQSHSDPYAHPPT 1264

QY 255 PRGGATQPKPP 267

Db 1265 PRFGISVPYSQPP 1277

RESULT 8  
ADE87987

ID ADE87987 standard; protein; 4019 AA.

XX ADE87987;

XX DT 29-JAN-2004 (first entry)

XX

DE Human lung tumour antigen polypeptide #92.  
 XX Human; lung tumour antigen; cancer; lung cancer; CD4+; CD8+; T cell;  
 KW immune response; immunostimulant; cytostatic.  
 XX Homo sapiens.  
 OS  
 XX US2003118599-A1.  
 FN  
 XX 26-JUN-2003.  
 XX  
 XX 10-MAY-2002; 2002US-00144649.  
 XX  
 XX 02-APR-1999; 99US-00285323.  
 PR 09-AUG-1999; 99US-00370838.  
 PR 30-DEC-1999; 99US-00476235.  
 PR 03-MAR-2000; 2000US-00518809.  
 PR 29-MAR-2000; 2000US-00538037.  
 PR 08-JUN-2000; 2000US-00588937.  
 PR 18-AUG-2000; 2000US-00640878.  
 PR 20-SEP-2000; 2000US-00667170.  
 PR 01-NOV-2000; 2000US-00704512.  
 PR 14-DEC-2000; 2000US-00738973.  
 PR 11-MAY-2001; 2001US-00854133.  
 XX (CORI-) CORIXA CORP.  
 PA  
 XX Algate PA, Lodes MJ, Wang T, Fan L, Mcneill PD;  
 PI  
 XX WPI; 2003-897103/82.  
 DR N-PSDB; ADE87984.  
 DR  
 XX New polynucleotides encode lung tumor antigens and are useful to  
 PT stimulate an immune response or detect or treat a cancer in a patient,  
 PT particularly lung cancer.  
 PT  
 XX Disclosure; SEQ ID NO 425; 63pp; English.  
 XX  
 XX The invention relates to polynucleotides encoding lung tumour antigens.  
 CC The invention also relates to the polypeptides encoded by the  
 CC polynucleotides, isolated antibodies or antigen-binding fragments that  
 CC specifically bind the polypeptides and a method for detecting cancer in a  
 CC patient, comprising obtaining a biological sample from the patient,  
 CC contacting the sample with a binding agent that binds a polypeptide of  
 CC the invention, detecting in the sample an amount of polypeptide that  
 CC binds to the binding agent, and comparing the amount of polypeptide to a  
 CC predetermined cut-off value. T cells specific for a tumour protein can be  
 CC stimulated and/or expanded by contacting the T cells with a polypeptide,  
 CC polynucleotide or an antigen-presenting cell that expresses a  
 CC polypeptide. Cancer development can be inhibited by incubating CD4+  
 CC and/or CD8+ T cells isolated from a patient with a polypeptide,  
 CC polynucleotide or an antigen-presenting cell that expresses a  
 CC polypeptide, so that the T cells proliferate. The invention is used to  
 CC stimulate an immune response or to detect or treat a cancer in a patient,  
 CC particularly lung cancer. This sequence represents a human lung tumour  
 CC antigen polypeptide of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification but was obtained in  
 CC electronic format from USPTO at seqdata.uspto.gov/sequence.html.  
 XX  
 XX Sequence 4019 AA;  
 SQ

Query Match 9.5%; Score 136.5; DB 7; Length 4019;  
 Best Local Similarity 24.0%; Pred. No. 0.053;  
 Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;  
 29 ARTSVVDETLFGSPAGTR-----PTPDDFPDPVVEK-----ANTRGVGX 68  
 988 SPLQVNETNARPQVRLDLCSSSTNNDPAKPDTPRPVMTQFPKSLGSLSPVVSE 1047  
 69 EASKALGAKGSC-----ETTPSRGSTPLTPRKXNK-----YRPISTPS 108  
 1048 QVAKGPFAAGTSDHFTKFSRADVFQRIQIPDSYARPLLTAPLDGSGPPEKTPMQPPS 1107

QY 109 YCDESLFGSRSEGASFGAPRMAGKAAKLALLWTPPTPEGSH--SPRPREAPLRAHP 166  
 DB 1108 SQDP--YGSVSQ-----ASRLSVDYERFAL--TPRPIDFNSHQNDYPSQPLTFHP 1158  
 QY 167 -----AGPSK--TEPG-----PAADSQKLSMGLHS-----SRPL 194  
 DB 1159 AVNESFAHPSRAFSQPGTISRPTSQDPYSQPPGTPRPVWDSYQSQSGTARSNTDPYSQP- 1217  
 QY 195 KRGLSHSLTHLVNSTGCHPATSAHTNGPDLPSTSGVTFRSPLYTSRARSYSISVPST 254  
 DB 1218 -----EGTPRPITVDYPSQQPTPRPSTOTDLFTVPTNQRHSDPYAHPPT 1264  
 QY 255 PRGGATQKPKPP 267  
 DB 1265 PRGISVPSQPP 1277  
 RESULT 9  
 ABP69736  
 ID ABP69736 standard; protein; 4025 AA.  
 XX AC ABP69736;  
 XX DT 20-JAN-2003 (first entry)  
 XX DE Human polypeptide SEQ ID NO 1783.  
 XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 KW cell-proliferative disorder; neurodegenerative disease; bacteria;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritic.  
 XX OS Homo sapiens.  
 XX FN WO200270539-A2.  
 XX PD 12-SEP-2002.  
 XX PF 05-MAR-2002; 2002WO-US005095.  
 XX PR 05-MAR-2001; 2001US-00799451.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Dmanac RT;  
 XX WPI; 2002-759812/82.  
 DR N-PSDB; ABZ11953.  
 XX New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
 PT or coagulation disorders.  
 XX  
 PS Claim 9; SEQ ID NO 1783; 1012pp + Sequence Listing; English.  
 XX The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleotide sequence selected from any of 948 sequences (ABZ1119-  
 CC ABZ12066) or their mature protein coding portion, active domain coding  
 CC protein or complementary sequences. The polynucleotides are useful for  
 CC identifying expressed genes or for physical mapping of human genome. The  
 CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
 CC markers, as a food supplement, for generating antibodies, in medical  
 CC imaging, screening and diagnostic assays and for treating cell-  
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,

CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 4025 AA;

Query Match 9.5%; Score 136.5; DB 5; Length 4025;  
Best Local Similarity 24.0%; Pred. No. 0.053;  
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPAGTR-----PTPDDPPVWEK-----ANRTRGVCK 68  
Db 994 SRPLQWNETTANRSPVRDLCSSTTNDPYAKPDPTRPVMTDQPKSLGSRSPVSE 1053  
QY 69 EASKALGAGKSC-----ETTPSRGSTPTLTTPRKNK-----YRPISHTPS 108  
Db 1054 QTAGPIAAGTSDHFTKPSRADVFORQIPDSYARPLTAPLDSGPGFPTKQPPPS 1113  
QY 109 YCDESLFGSRSEGAFCAPRMAKDAKRALLTWPTTPRGSH--SPRPREAPLRAIHP 166  
Db 1114 SQDP--YGVSVQ-----ASRRLSVDPYERPAL--TPRPIDNFSHNSQNDPYQPPLTPHP 1164  
QY 167 -----AGPSK--TEPG-----PAADSQKLSMGGLHS-----SRPL 194  
Db 1165 AVNESFAHPSRAFSQPGT--SRPTSQDPYQPGTGPTRPVVDSYQSSGTRARSNTDPSQD- 1223  
QY 195 KRGLSHSLTLNVPSTGCHATSAPHNTGPDRLRPSTGVTFRSPLVTSRARSVS--SVPT 254  
Db 1224 -----PGTRPTTVDPSQQTFRPSTQTLFVTPVNRHSDDYAHPPGT 1270  
QY 255 PRGGATQPKPP 267  
Db 1271 PRPGISVPYQPP 1283

RESULT 10

ID ABB63299 standard; protein; 2703 AA.  
XX  
AC ABB63299;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 16689.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL07402.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 16689; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA  
CC sequences (ABU1840-ABU16175) and the encoded proteins (ABBS7737-  
CC ABBS72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2703 AA;

Query Match 9.5%; Score 136; DB 4; Length 2703;  
Best Local Similarity 22.5%; Pred. No. 0.036;  
Matches 62; Conservative 30; Mismatches 119; Indels 64; Gaps 11;

QY 23 GGYRV-----KARTSYVDETLFGSPAGTRTPPDF-----DPPWVEKANRTRGVG 67  
Db 633 GGYKMGPGGPGAGQYPPQPOQYPPGNYPYPPQYPPGAYATGPPPPPTSQA-----GAG 688  
QY 68 KEASKALGAKGSCETTPSRGSTPTLTTPRKNKYRPISHITPSYCDSELFGRSEGAFCAP 127  
Db 689 GANSMPGSAQAG--GYFGRG-----MENHTGQYPPYQWVPPSPQOTVPGGAPGAMVGNH 741  
QY 128 RMAKGDAAKRALLTWP-----PPTPRGSHSPREAPLRAIHPAGPSKTEPGPAADS 180  
Db 742 VQGRG-----TPPPVVGPPPPPGSGSPRLNLYLKHQLOHKGVGSGSTPPQGP 791  
QY 181 QKLSMG--GLHSSRPL-----KRGLSHSLTLNVPST-----CHPATSAFHT 220  
Db 792 QGYNGPTGMFGPMGPMGPHMGPHGFTNMGPTSTPQSQMLQGGPQOQSGAGGPES 851  
QY 221 NGPDRLRPSTGVTFRSPLVTSRARSVSISVPSTP 255  
Db 852 GGPEHI--SQDNGISSSGTGAAGMHAVTSSVVTGP 885

RESULT 11

ID ABB70019 standard; protein; 2703 AA.  
XX  
AC ABB70019;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE Larval viability associated protein #18.  
XX  
KW Fruit fly; larval viability; insecticidal activity; maize; wheat; oat;  
KW rye; sorghum; rice; barley; millet; turf; cotton; sugarcane; sugar beet;  
KW oilseed rape; soybean; vegetable crop; fruit.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200257455-A2.  
XX  
PD 25-JUL-2002.  
XX  
PF 18-JAN-2002; 2002WO-US001568.  
XX  
PR 18-JAN-2001; 2001US-0262351P.  
XX  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Stam L, Bachmann J, Broadus J, Kamdar KP;  
PI  
XX WPI; 2002-590746/63.  
DR N-PSDB; ABSS1395.  
XX  
XX Identifying inhibitors of activity of proteins essential for Drosophila  
PT larval viability comprises expressing in a host a protein essential for  
PT larval activity and identifying compounds that inhibit or interact with

PT the protein.  
 XX Claim 1; Page 108-117; 169pp; English.  
 XX  
 CC The invention describes a method of identifying compounds that inhibit  
 CC the activity of, or that interact with a protein essential for Drosophila  
 CC larval viability comprising expressing in a recombinant host a DNA  
 CC molecule to produce a protein essential for larval viability. The method  
 CC is useful for identifying compounds with insecticidal activity. Compounds  
 CC identified are useful as insecticides in crops such as maize, wheat,  
 CC oats, rye, sorghum, rice, barley, millet, turf, cotton, sugarcane, sugar  
 CC beet, oilseed rape, soybeans, vegetable crops and fruits. This is the  
 CC amino acid sequence of a fruit fly larval viability associated protein  
 XX  
 XX Sequence 2703 AA;  
 Query Match 9.5%; Score 136; DB 5; Length 2703;  
 Best Local Similarity 22.5%; Pred. No. 0.036;  
 Matches 62; Conservative 30; Mismatches 119; Indels 64; Gaps 11;  
 QY 23 GGYRV-----KARTSYVDLTFGSPAGTPTPDP-----DPPNVEKANTRGVG 67  
 DB 633 GGYKMGFGGSGAQGPPQPOQYFNGYPPRPQPGAYATGPPPPPTSQA-----GAG 688  
 QY 68 KEASKALGAKGSCETTPSRGSTTLTPRKKNKYRPISTPSYCDLSFGSRSEGASFGAP 127  
 DB 689 GANSMPSGAQAG--GYFGRG-----NENHTGQYPPYQWVPPSPQQTVPFGAGPGAMVGNH 741  
 QY 128 RMAKGAADAKLALLWTP-----PPTPGSHSPRPREAPLRAIHPAGSKTEPGPAADS 180  
 DB 742 VQKGK-----TPPPVVGPPGPPQPGSGSPRELNYLKQHLQHKGYGSGSPTPPQGP 791  
 QY 181 QKLSMG--GLHSSRPL-----KRGLSHSLTLNVST-----GHPATSAPHT 220  
 DB 792 QGYGNGPTGHPGMPGPHMGPPHMGPPHMGPPHMGPPHMGPPHMGPPHMGPPHMGPPH 255  
 QY 221 NGQDLRPSTGTVFRSPLVTSRARSVSISVPSTP 255  
 DB 852 GGPEHI--SQDNGISSSGPTGAAGMHAIVTSVITGP 885  
 RESULT 12  
 ABP76681  
 ID ABP76681 standard; protein; 19938 AA.  
 AC  
 XX ABP76681;  
 XX  
 DT 26-FEB-2003 (first entry)  
 XX  
 DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 5.  
 XX  
 KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
 KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
 XX  
 OS Streptomyces viridochromogenes.  
 XX  
 PN WO200268436-A1.  
 XX  
 XX 06-SEP-2002.  
 XX  
 XX 24-AUG-2001; 2001WO-BP009815.  
 XX  
 XX 25-FEB-2001; 2001DB-01009166.  
 XX  
 XX (COMB-) COMBINATURE BIOPHARM AG.  
 XX  
 XX Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
 XX WPI; 2003-018650/01.  
 XX N-PSDB; AB237516.  
 XX  
 XX New avilamycin derivatives, useful for treatment of infections, and  
 XX nucleic acid encoding avilamycin synthesis enzymes.  
 PT

XX Example 1; Page 68-301; 319pp; German.  
 XX  
 CC The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoacide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
 CC viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-  
 CC AB237516)  
 XX  
 XX Sequence 19938 AA;  
 Query Match 9.4%; Score 134.5; DB 6; Length 19938;  
 Best Local Similarity 26.8%; Pred. No. 0.55;  
 Matches 75; Conservative 22; Mismatches 94; Indels 89; Gaps 15;  
 QY 49 PPDPDPWVEKANTRGVGKEASKALGAKGSC-----ETTPSR 86  
 DB 11439 PSELPOQWVE---RXROSETPEPMSASSQRCRTWXRSPAQSVASKWPSPANSDTSASR 11495  
 QY 87 GSTTTLTP-----RKKNKYRP-----ISHTPSYCDLSFGSRSEGASFGAPRMAKGA 134  
 DB 11496 SST-LVTPTCTSMNTNSARRFRSPSRARIS-TPG-CPDIARVSAAPSAPSSRRLAPGNA 11552  
 QY 135 AK-LRALLW-----TPPPTPRGSH--SPRPREAPLRAIHPAGSKTEPG 175  
 DB 11553 AKNASASWGWLSRKSRRVEAVRYRTPLYTKSFTHGSSSTRRRISRSFTVALP---PL 11608  
 QY 176 PAADSQKLSGGLHSSRPLKRGLSHSLTLNVSTGHPATSAP-----HTNGPQD- 225  
 DB 11609 SGALPSETAWAGDHRSAFLRSATH-----CAHSGTRSPPIISKEERSAGPTEA 11659  
 QY 226 -----LRPSTGTVFRSPLVTSRARSVSISVPSTPRRG 259  
 DB 11660 AHAARASPSASGVSAQSP-KTSRASSVAVXATITTSRAG 11698  
 RESULT 13  
 ADA24266  
 ID ADA24266 standard; protein; 502 AA.  
 XX  
 AC ADA24266;  
 XX  
 XX 20-NOV-2003 (first entry)  
 XX  
 DE Human WASP interacting protein (WIP) amino acid sequence.  
 XX  
 KW WASP interacting protein; WIP; transgenic; Wiskott-Aldrich syndrome; WAS;  
 KW x-linked immunodeficiency; WAS protein; WASP; thrombocytopenia; eczema;  
 KW impaired immunity; lymphoma; leukaemia; WIP agonist; WIP antagonist;  
 KW cytoskeletal; T cell receptor; T cell activation; B cell proliferation;  
 KW human.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 XX Region 8..13  
 FT /label= APPPPP\_motif  
 FT Misc-difference 231..232  
 FT /note= "Encoded by GGAGGAGGC"  
 FT Misc-difference 359..360  
 FT /note= "Encoded by CCAGTGCCC"  
 FT Region 426..431  
 FT /label= APPPPP\_motif  
 XX  
 XX WO2003070893-A2.  
 XX  
 XX 28-AUG-2003.  
 XX  
 XX 18-FEB-2003; 2003WO-US004646.  
 XX  
 XX

```
PR 19-FEB-2002; 2002US-00078547.
PA (CHIL-) CHILDRENS MEDICAL CENT.
XX Ramesh N, De La Fuente MA, Anton IM, Geha RS;
PI WPI; 2003-712614/67.
XX DR N-PSDB; ADA24265.
XX
XX New transgenic non-human mammal, useful for enhancing or inhibiting T
PT cell receptor-mediated T cell activation or B cell proliferation in a
PT mammal for treating or preventing diseases, e.g., leukemia.
XX
XX Example 1; Fig 1D; 106pp; English.
PS
XX This invention relates to a novel WASP interacting protein (WIP) and the
CC DNA sequence which encodes it. In addition, the invention relates to a
CC novel transgenic non-human mammal whose genome comprises a disruption of
CC a WIP gene. Wiskott-Aldrich syndrome (WAS) is an X-linked
CC immunodeficiency caused by mutations that affect the WAS protein (WASP).
CC It is characterised by thrombocytopenia, eczema, impaired immunity and a
CC predisposition to develop lymphomas and leukaemias. Modulation of the
CC activity of a WIP protein, through use of a WIP agonist or WIP
CC antagonist, may produce cytostatic activity. The transgenic non-human
CC mammal may be useful for enhancing or inhibiting T cell receptor-mediated
CC T cell activation or B cell proliferation in a mammal for treating or
CC preventing diseases such as leukaemia. The present sequence is the amino
CC acid sequence of the human WIP protein of the invention.
XX
XX Sequence 502 AA;
Query Match 9.3%; Score 134; DB 6; Length 502;
Best Local Similarity 25.3%; Pred. No. 0.0067;
Matches 66; Conservative 22; Mismatches 89; Indels 82; Gaps 13;
QY 42 PAGTR-----PTPPDFPPVWEKANTRGVGKEASKALGAKGSCETTPRGSTP----- 90
DB 207 PGGPRQSPGPTTPTT-----PQNRGALG-----GSIROSPLSSSPFENRPP 250
QY 91 -----TLTPRKKNKYRTPISHTPSYCDESLFGSRSEGASFGAPRMAGDAKLRALLWT 143
DB 251 LPTTPSRALDKPPPPPPVGNRPSTHREAV----- 281
QY 144 PPPTPRGSHSPRPR-----EAPLRA-IHPAGPSKTEPGPAADSQKLSMGGLHSRPLKRL 198
DB 282 PPPPPQNNKVPSTPRPSAPRPHLEPPPPSPRPPPLPPS---SSGNDETETRLPQNL 338
QY 199 SHSLTHLNVSTGH--PATSAPHTNGQDLR--PSTSGVTFRSPLVTSRARSVISVPST 254
DB 339 SLSSSTPLPSGRSGPLPPPPGPPPPVDPGRSG-PLPPPPPVSRNGSTSRALPAT 397
QY 255 PR---RGATQK---PKPP 267
DB 398 PQLPSRSGVDSRSGPRPP 416
RESULT 14
ID ABG21932 standard; protein; 470 AA.
XX
AC ABG21932;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #1923.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
EN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS86119.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 52291; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 470 AA;
Query Match 9.2%; Score 132.5; DB 4; Length 470;
Best Local Similarity 25.1%; Pred. No. 0.0083;
Matches 66; Conservative 18; Mismatches 94; Indels 85; Gaps 12;
QY 42 PAGTRTPPPDFPPVWEKANTRGVGKEASKALGAKGSCETTPRGSTPTLTPRKKNKYR 101
DB 240 PPSATPSPP-----TGSPATGRTSFRTELATPRR-----R 269
QY 102 PISHTPSYCDESLFGSRSEGASFGAPRMAGDAKLRALLWTTPP-----TPRGSH 152
DB 270 AASQADTPCPC---GTPSRAP--SPR-TRPPAARFAPXAWPRPPSVAGARWWTXRGAP 323
QY 153 SP-----RPREAPLRAIHPAG---PSKTEPGPAADSQKLSMGGLHSRPLKGLSHS 201
DB 324 APPGPPGGRPARLPRAAGPAGRRHRPDRGPPSPAQRDPRAAPRTSRSASR----- 378
QY 202 LTHLNVSTGHPATSA-----PHT-----NGPQDLRSTSGVTFRSPLVTSRARS 246
DB 379 -----APRSAGFAAAGTATAARPPATQARAPRAPRRPRPATSSXTARAPPPSGAKS 433
QY 247 VSISVFSTPRRGATQKPKPPWK 269
DB 434 AE-----PPPEQPPPPPPPPAPPKK 452
RESULT 15
ID AAM39328 standard; protein; 705 AA.
```

XX AAM39328;  
 XX 22-OCT-2001 (first entry)  
 XX Human polypeptide SEQ ID NO 2473.  
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX Homo sapiens.  
 XX WO200153312-A1.  
 XX 26-JUL-2001.  
 XX 26-DEC-2000; 2000WO-US034263.  
 XX 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-0052317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 XX WPI; 2001-442253/47.  
 DR N-PSDB; AAI58484.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX Example 4; SEQ ID NO 2473; 10078pp; English.  
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 XX Sequence 705 AA;  
 XX  
 XX Query Match 9.2%; Score 131.5; DB 4; Length 705;  
 XX Best Local Similarity 25.2%; Pred. No. 0.017;  
 XX Matches 61; Conservative 18; Mismatches 110; Indels 53; Gaps 8;  
 XX  
 XX 40 GSPAGTRPTPPDPPWEKANTRG-----VGKEASKALGAKGSCETTSRGSTP 90  
 XX  
 XX 454 GPPAQQRPPQGPPQPGFQRPQPPQPGQGHLSGLGPPAG-SPLPQRLPSP 511  
 XX  
 XX 91 TLTPRKKNYRPISTHPSYCDSELSFGSRSEGAFCGAPRMAGDAAKRLALLWTPPTPR- 149

Db 512 TSAPO-----QFASQAAPPTQGRQSRPVAGGFCAPPFAARPPASPPQRAQPPQATRQ 566  
 QY 150 -----GSHSPRPRAEPLRAIHGAPGSKTEPGPAADSQKLSMGLHSSRPLKRGLSHLTHL 205  
 Db 567 TSVSGPAPPKASGAPPGGQQQPGPPQKPPGAGPTRQASQAG----- 608  
 QY 206 NVPSTGHPTASAPHTNGQDLRPSTSGVTFPSPLVTSRARSVSISVSPTRRGATQPK 265  
 Db 609 PVPRTGPPTTQQP-----RPSGPGPAGRPKPKQLAQKFSQDVPPPTATAAAG----- 654  
 QY 266 PP 267  
 Db 655 PP 656  
 Search completed: August 10, 2004, 12:11:12  
 Job time : 56 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 12:11:17 ; Search time 19 seconds  
(without alignments)  
730.915 Million cell updates/sec

Title: US-10-031-589-4  
Perfect score: 1437  
Sequence: 1 MKTPVELAVSGMQLGLQHR.....SVPSTPRRGATQKPKPWK 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147.5	10.3	1706	4	US-09-252-991A-31760
2	135.5	9.4	920	4	US-09-252-991A-28918
3	134.5	9.4	503	4	US-09-599-287A-2
4	132.5	9.2	1008	4	US-09-352-991A-29419
5	130.5	9.1	878	4	US-09-556-706B-2
6	128.5	8.9	501	4	US-09-252-991A-17730
7	128.5	8.9	536	4	US-09-252-991A-16754
8	128	8.9	507	4	US-09-599-287A-24
9	126.5	8.8	693	4	US-09-252-991A-26071
10	126	8.8	696	3	US-08-906-865-4
11	126	8.8	696	4	US-08-129-668-4
12	124.5	8.7	1034	4	US-09-252-991A-28921
13	123.5	8.6	315	4	US-09-252-991A-16743
14	123.5	8.6	390	4	US-09-252-991A-17829
15	123.5	8.6	1321	2	US-08-317-310A-64
16	122.5	8.5	1037	4	US-09-252-991A-17548
17	122	8.5	428	4	US-09-252-991A-25955
18	122	8.5	904	4	US-09-976-594-515
19	121.5	8.5	309	4	US-09-252-991A-25386
20	121.5	8.5	907	3	US-08-783-774-2
21	121.5	8.5	907	4	US-08-328-599A-1
22	121.5	8.5	907	5	PCT-US95-04611A-19
23	120.5	8.4	202	4	US-09-252-991A-32054
24	120	8.4	345	4	US-09-252-991A-18076
25	120	8.4	405	4	US-09-252-991A-27573
26	119.5	8.3	189	4	US-09-252-991A-16638
27	118	8.2	315	4	US-09-252-991A-20553

28	118	8.2	663	4	US-09-252-991A-30843	Sequence 30843, A
29	117.5	8.2	432	4	US-09-252-991A-30848	Sequence 30848, A
30	117.5	8.2	906	4	US-09-252-991A-32715	Sequence 32715, A
31	117	8.1	658	4	US-09-328-599A-2	Sequence 2, Appli
32	116.5	8.1	793	4	US-09-252-991A-18035	Sequence 18035, A
33	116.5	8.1	312	4	US-09-252-991A-30114	Sequence 30114, A
34	116	8.1	681	4	US-09-688-188B-29	Sequence 29, Appli
35	116	8.1	681	4	US-09-765-815-2	Sequence 2, Appli
36	116	8.1	681	4	US-09-291-417D-29	Sequence 29, Appli
37	115.5	8.0	280	4	US-09-252-991A-19400	Sequence 19400, A
38	115.5	8.0	387	4	US-09-489-847-372	Sequence 372, App
39	115	8.0	548	4	US-09-252-991A-19915	Sequence 19915, A
40	114	7.9	495	4	US-09-252-991A-31949	Sequence 31949, A
41	114	7.9	571	4	US-09-134-001C-3865	Sequence 3865, A
42	114	7.9	1046	4	US-09-252-991A-27508	Sequence 27508, A
43	114	7.9	1162	2	US-08-728-323A-2	Sequence 2, Appli
44	114	7.9	1162	4	US-09-298-568-2	Sequence 2, Appli
45	114	7.9	1162	4	US-09-410-399-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-252-991A-31760  
; Sequence 31760, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107156.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31760  
; LENGTH: 1706  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31760

Query Match	10.3%	Score 147.5	DB 4	Length 1706
Best Local Similarity	24.7%	Pred. No. 0.00011		
Matches	64	Conservative 21	Mismatches 103	Indels 71
Gaps	9			
QY	19	HRCRGYRVKARTSYVDTELFSGPAGTRPTPPDFDPWVEKANTRGVGKESKALGAKG	78	
Db	1277	HARRASLHALART-----GTLAGRATTGEDRKRIPOARPRTR-----	1314	
QY	79	SCETTPSRGSTPTLTPRKKNKVRPI--SHTPSYCDSEIFGSRSEGSFGAPRMKGDAAK	136	
Db	1315	RARHPFGRGITPATPPPGKQRPAPGPRRQRCQR--GSPANPASAGRR-----	1364	
QY	137	LRALLWTPTTPTRGSHSP-----RPEAPLRAHPAGSKTEPPGPAADSQKLSMGLHSSR	192	
Db	1365	-----PFRRGSPAPRPFQRPQRRTGTPQSPARTPAPED-----	1402	
QY	193	PLKRGSLSLTHLNVPTSGHPATSPHTNGPDLPSPSTGVTFRRSPLV-----TSRARSV	247	
Db	1403	--RRNAHPGTIVARPPATGRPARTAGAHRAHRTATARAARRGQRPSPRATGTRSR--	1458	
QY	248	SISVSTPRRGATQKPKP 266		
Db	1459	--TAPAGOPAAARQGP RP 1475		

RESULT 2  
US-09-252-991A-28918  
; Sequence 28918, Application US/09252991A

Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 28918  
 LENGTH: 920  
 TYPE: PRT  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-28918

Query Match 9.4%; Score 135.5; DB 4; Length 920;  
 Best Local Similarity 26.3%; Pred. No. 0.0061;  
 Matches 78; Conservative 24; Mismatches 110; Indels 85; Gaps 16;  
 QY 5 VELAVSQMQLGLQHCRCGGRVYKARTSYVDETLF---GSPAGTRTPDPDFPPWYERAN 61  
 DB 651 VLRLVRLQQQGVVHRDVLGLAIAAGHEY---LFAVEGQFQHOR--HPESD----- 696  
 QY 62 RTRGVGKEASKALGAKGSCETTPTLTPRKNKYRPISHTSYCDESLFGSR--- 118  
 DB 697 RQGGFGLQAR-LQMLVSGRRVCHRAPASATSTGRCRPPGRP--CRPRSAGNRPR 753  
 QY 119 -SEGASGAPRMA-----KGDAKRLALLWTPPTPR-----GSHSRP----- 156  
 DB 754 APPCAMAGARRRTRGHRGSGASAGW-PPAPAADRGCRAGGAGSHSGVHSAPG 812  
 QY 157 --REAPLRAHPAGSKTEPGPAADSKLMSGLHSSRLKGLSHSLTHLNVSTGHPA 214  
 DB 813 AGARPLPWSGEPANPAGPA-----PCARRR-----A 845  
 QY 215 TSAPHTN---GPDLRPTSGVTFRPLVTSRARSVSISVPSTPRRGGATKPKPPW 268  
 DB 846 TASCHRHAFGWEDAGLPGAGCS-RSTATTDA-----PPASPRSGAWCGSRPAW 895

## RESULT 3

US-09-599-287A-2  
 Sequence 2, Application US/09599287A  
 Patent No. 6635446  
 GENERAL INFORMATION:  
 APPLICANT: Narayanaswamy Ramesh  
 APPLICANT: Ines M. Anton  
 APPLICANT: John H. Hartwig  
 APPLICANT: Raif S. Geha  
 TITLE OF INVENTION: WIP, A WASP-Associated Protein  
 FILE REFERENCE: 1242.1022-004  
 CURRENT APPLICATION NUMBER: US/09/599,287A  
 CURRENT FILING DATE: 2000-06-22  
 PRIOR APPLICATION NUMBER: PCT/US98/27501  
 PRIOR FILING DATE: 1998-12-22  
 PRIOR APPLICATION NUMBER: 60/101,457  
 PRIOR FILING DATE: 1998-09-23  
 PRIOR APPLICATION NUMBER: 60/068,533  
 PRIOR FILING DATE: 1997-12-23  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 503  
 TYPE: PRT  
 ORGANISM: Human  
 US-09-599-287A-2

Query Match 9.4%; Score 134.5; DB 4; Length 503;  
 Best Local Similarity 25.8%; Pred. No. 0.00035;

Matches 63; Conservative 23; Mismatches 107; Indels 51; Gaps 11;  
 QY 42 PAGTR-PTPPDPDFPPWVEKANRTRGVGKEASKALGAKGSCETTPTSGSTPT--LTPRKN 98  
 DB 207 FGGPQPPGPTTPTTFCNRTALGGSIROSPLSSSPFSNRPLPTTSRALLDDKPP 266  
 QY 99 KYRPISTPTSCDSLSLFGSRSEGASFGAPRMVAKDAKRLALLWTPPTPRGSHSPRR- 157  
 DB 267 PPPPVGNRPSIHRFAV-----PPPPQNNKPPVPT 297  
 QY 158 ---EAPLEA-IHPAGPSKTEPGPAADSKLMSGLHSSRLKGLSHSLTHLNVSTGH- 212  
 DB 298 PRPSAPHRPHLRPPPPSRGPPPLPPS---SSGNDETFRLPQRLNLSLSSSTPPLSPGRS 354  
 QY 213 -PATSAPHTNGQDLR--PSTSGVTFRPLVTSRARSVSISVPSTPR---RGGATQK--- 263  
 DB 355 GPLPPPPSERPPFVRDPPGRSG-PLPPPPPVSRNGSTSRALPATPOLPSRSGVDSPRSG 413  
 QY 264 EKPP 267  
 DB 414 PRPP 417

## RESULT 4

US-09-252-991A-29419  
 Sequence 29419, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 29419  
 LENGTH: 1008  
 TYPE: PRT  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-29419

Query Match 9.2%; Score 132.5; DB 4; Length 1008;  
 Best Local Similarity 24.8%; Pred. No. 0.0013;  
 Matches 76; Conservative 25; Mismatches 141; Indels 65; Gaps 13;  
 QY 17 LQHCRCGGYR---VKARTSYVDETLFGSP-AGTRTPPTDFPPWVEKANRTRGVGKEASK 72  
 DB 377 LVRRGGRSRPPVVRAGTGHGRRRPAQQAAGADRT---DPPQSGPASSARPPARRR 432  
 QY 73 ALGAKGSCETTPSRGSTPTLTPRKK---NKYRPISTPTSCDSLSLFGSRSEGASFGAPRM 129  
 DB 433 ATGAAGTQPAAGA--ALRPEAGAGGAFRLPGRPTAAQAAPGRTGAGPGRSRV 490  
 QY 130 AKGDAKRLALLWTPPTT--PRGSHSPRPREAPLRAHPAGPSKTEPGPAADSKLMSGL 187  
 DB 491 AAGLGRRRQPARLRPPPTQFPGSSGDPAAARPARRTAALPARPELDADPARTQRRRAGR 550  
 QY 188 LHSS-----RPLKGLSHSLTHLNVSTGH-----PATSAPTN 221  
 DB 551 RHGAGQDLAVAGRAPCKAG--RAARHAGAGGADAHQDPDLAGRGRALRPRSRAGARR 608  
 QY 222 GPQD-LRP---STSGVTFRPLVTSRARSVSISV-PSTPRRG-----G 259  
 DB 609 RPAPRLRPHRRARPAGADHLCPAATRRRRAQAQAVFPADPRRGAERHQRHQRPRRPRAG 668  
 QY 260 ATQKPK 266  
 DB 669 GTASPPV 675



```
Db 366 -RALAGCPVPAAGRCGWSRPARSGSP---RPPVAG---RDSPGRRAGAGSVADSAPG 418
QY 254 T-----PRGG-----ATQKPKP 266
Db 419 TRRARAAGCGRSRRAKAAVPRRSRGAATGAPPP 456

RESULT 8
US-09-599-287A-24
; Sequence 24, Application US/09599287A
; Patent No. 6635446
; GENERAL INFORMATION:
; APPLICANT: Narayanaswamy Ramesh
; APPLICANT: Ines M. Anton
; APPLICANT: John H. Hartwig
; APPLICANT: Raif S. Geha
; TITLE OF INVENTION: WIP A WASP-Associated Protein
; FILE REFERENCE: 1242.1022-004
; CURRENT APPLICATION NUMBER: US/09/599,287A
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/068,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translated WIP ORF No. 6635446 3
US-09-599-287A-24

Query Match 8.9%; Score 128; DB 4; Length 507;
Best Local Similarity 25.3%; Pred. No. 0.0014;
Matches 62; Conservative 23; Mismatches 108; Indels 52; Gaps 11;

QY 42 PACTR-PTPDPDPVWEKANTRGVGKEASKALGAKGSCETTPSRGSPPT--LTPRKKN 98
Db 210 PGGPRQPSGPTPPPPGNGRTALGGSIHQPLSSSSPFSNRPLPPTPSALDDKPPP 269
QY 99 KYRPISHTSYCDSELFGRSEGCASFAPMAKGDAAKURLLWTPPTTPRGSHSPRR- 157
Db 270 PPPVGNRPSIHREAV-----PPPPQNNKPPVPST 300
QY 158 ---EAPLRA-IHPAGSKTEPGPAADSQKLSMGLHSSRLPKRGLSHSLTHLVNVPSTGHP 213
Db 301 PRSAHRHLRPPPSRPGPPLPS---SSGNDETPLPORNLSLSSTPLPSGRS 357
QY 214 A---TSAPHTNGQDLR--PSTSGVTFRSLVTSRARSVISVPSTPR---RGATQK-- 263
Db 358 GPLPFPVPSRPPPPVDPDPGRSG-PLPPPPVPSRNGSTSRALPATPQLPSRSGVDSPRS 416
QY 264 -PKPP 267
Db 417 GPRPP 421

RESULT 9
US-09-252-991A-26071
; Sequence 26071, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
```

```
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26071
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26071

Query Match 8.8%; Score 126.5; DB 4; Length 693;
Best Local Similarity 26.9%; Pred. No. 0.0028;
Matches 87; Conservative 28; Mismatches 112; Indels 97; Gaps 19;

QY 16 GLQHRCR-----GGVYVKARTSYVDETLFGSP---AGTRPTPPDFDPWVEKANTR 64
Db 332 GLPQARQPAATAGCPGTAGRRSGVADP--GGPLOGAGRRLEP-----NVRP 378
QY 65 GVGKEASKALGAKGSCETTPSRGSTPTLTP-----RKNKYRPISHTSYCDSELFGRSE 120
Db 379 GVGLRRRRARTGEGVNGGPAAR--YPLRLPRGGARQLHPGPGAGDLS--GDFAPGRRAR 435
QY 121 CASFGAP-----RVAKGD---AAKLRLWTP-----PPTPRGSH 152
Db 436 TGALGAPAAATQYAAQRGDKRSALLRALQADPTHGRRRRGRRRQVPVPRQRSPAPPRGDR 495
QY 153 S-PPREAP-----LRAIHPAGPSKTEPGPA-----ADSOKLSMGLHSSRPL 194
Db 496 TGPEAPDAATGGLPGRSGRLRRPH-AGPG--HPGPARGPRGRADHPQPNLAGLDPRGR 552
QY 195 KR-GLSHSLTHLVNVPSTGHPATSAPTHNGP-----QDLRPSGTVTFRSLVTSRA 244
Db 553 SRPGVQRGLCRAGIPGTAWRAT---HAGGPAASLPERGRAVLPAVLGLRLRRTRRNAARS 609
QY 245 RVSISIVPS---TPREGGATQKP 264
Db 610 RRGAGQGPSRGLRRRRGRARRUP 633

RESULT 10
US-08-906-865-4
; Sequence 4, Application US/08906865
; Patent No. 6040168
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Porton, Barbara
; APPLICANT: Kao, Hung-Teh
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,865
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
```

TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 696 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: /desc = "Synapsin Ia"  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-906-865-4

Query Match 8.8%; Score 126; DB 3; Length 696;  
Best Local Similarity 24.1%; Pred. No. 0.0031;  
Matches 60; Conservative 21; Mismatches 114; Indels 54; Gaps 10;  
QY 40 GSPAGTRPTPPDPPWVEKANRTRG-----VGKEASKALGAKGSCETTPSRGSTP 90  
DB 454 GPPAQQEP-PPQGGPPQGGPQGGPPLQQRPPQGGHLSGLGPPAG-SPLPQRLPSP 511  
QY 91 TLTPRKKNKYRPISTHTSYCDSELFGRSEGASFGAPRMAGDAKRLALLTTPPTPR- 149  
DB 512 TSAFQ-----QPASQAAPPTQGGQSRPVAGGFGAPPAARPPASPPQAGPPQATRQ 566  
QY 150 -----GSHSPREAPLRAIHAPGSKTEPCGPAADSKLSMGLHSSRPLKRLGSLHSLTHL 205  
DB 567 TSVSGAPPKASGAPPGQQRQPPQKPGPAGPTQASQAG----- 608  
QY 206 NVPTSGHPATSHPTNGPDRLRSTSGVTFRSPLVTSRARSVISVPSTPRRGATQ--- 262  
DB 609 PVRTGPTTTPQPRPSG-----FGPAGAP--KPQLACKP--SQDVPPPPATAAAGGPPHPQL 660  
QY 263 ----KPKPP 267  
DB 661 FNLPEPAPP 669

RESULT 11  
US-09-129-668-4  
; Sequence 4, Application US/09129668B  
; Patent No. 6429010  
; GENERAL INFORMATION:  
; APPLICANT: Greengard, Paul  
; APPLICANT: Porton, Barbara  
; APPLICANT: Kao, Hung-Teh  
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE AND USES  
; FILE REFERENCE: 600-1-202 CIP  
; CURRENT APPLICATION NUMBER: US/09/129,668B  
; EARLIER FILING DATE: 1998-08-05  
; EARLIER FILING DATE: 1997-08-06  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 696  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-129-668-4

Query Match 8.8%; Score 126; DB 4; Length 696;  
Best Local Similarity 24.1%; Pred. No. 0.0031;  
Matches 60; Conservative 21; Mismatches 114; Indels 54; Gaps 10;  
QY 40 GSPAGTRPTPPDPPWVEKANRTRG-----VGKEASKALGAKGSCETTPSRGSTP 90  
DB 454 GPPAQQEP-PPQGGPPQGGPQGGPPLQQRPPQGGHLSGLGPPAG-SPLPQRLPSP 511  
QY 91 TLTPRKKNKYRPISTHTSYCDSELFGRSEGASFGAPRMAGDAKRLALLTTPPTPR- 149  
DB 512 TSAFQ-----QPASQAAPPTQGGQSRPVAGGFGAPPAARPPASPPQAGPPQATRQ 566

QY 150 ----GSHSPREAPLRAIHAPGSKTEPCGPAADSKLSMGLHSSRPLKRLGSLHSLTHL 205  
DB 567 TSVSGAPPKASGAPPGQQRQPPQKPGPAGPTQASQAG----- 608  
QY 206 NVPTSGHPATSHPTNGPDRLRSTSGVTFRSPLVTSRARSVISVPSTPRRGATQ--- 262  
DB 609 PVRTGPTTTPQPRPSG-----FGPAGAP--KPQLACKP--SQDVPPPPATAAAGGPPHPQL 660  
QY 263 ----KPKPP 267  
DB 661 FNLPEPAPP 669

RESULT 12  
US-09-252-991A-28921  
; Sequence 28921, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28921  
; LENGTH: 1034  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28921

Query Match 8.7%; Score 124.5; DB 4; Length 1034;  
Best Local Similarity 24.2%; Pred. No. 0.0071;  
Matches 80; Conservative 25; Mismatches 114; Indels 111; Gaps 16;  
QY 42 PAGTRPTP-----PDFDFPWV-----EKANRTR-----GVGKE 69  
DB 2 PAGDPPADSRDACHRPVADPPGAGPAQAARPAHQARDRDRRLRRQRAYPGAAGVGRP 61  
QY 70 ASKALGAK-----GSCETTPSRGSTPTLTPRKKNKYRPISTHTSYCDSELFGRSEG--- 121  
DB 62 AHRAQPARPETFDHARARPARPRRPAAGPAGTPGPPPPAADAARERTRAPGHPQ 121  
QY 122 ---ASFGAPRMAGDAKRL---ALLWTPPTTPRGSHSP---RPREAPL----- 161  
DB 122 RPRAPCGAPRRARRRHAEGSCAGGWQLPQGRPHPPQAARRKBAAPLRGDRRLRPA 181  
QY 162 -----RAIHAPGSKTEPG-----PAADSQKLSMGLHSS-----RPLKRL 198  
DB 182 PAARPRPRQPRPAGPRRHAGGRSRERPAAPA-ALGRFGLLAAPAAAPAGRLCRTLAV 240  
QY 199 SHSLTHLNVFSTGH-----PATSHPTNGPDRLRSTSGVTFRS-PLVTSRAR 245  
DB 241 LPAAVDABPAATGHLGVRRPLAASGPARPAPVATGP-----GGRGGEALRGLPL-----ER 292  
QY 246 SVSISVPSTPRRGATQKFKP-----PMK 269  
DB 293 PVALPGPRVRRGPGRLPRRPPGRSLQPWQ 322

RESULT 13  
US-09-252-991A-16743  
; Sequence 16743, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16743
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16743

Query Match
Best Local Similarity 24.1%; Score 123.5; DB 4; Length 315;
Matches 76; Conservative 29; Mismatches 100; Indels 111; Gaps 17;

QY 24 GYRVKARTSYVDETLFG--SPAGTRPTPPDFDPWVEKANTRGVGKKEASKALGAKGSC 81
DB 17 GLRRPRAAATQPRAGRAPGGRRLPREL-----ATGLGPGAGRGWVVRGARP 66
QY 82 TTPSRGSTPTLTPRKKNKYRPISHTPSYCYDESIFGSRSE--GASFGAPR-----MAKGA 134
DB 67 APRLPGAAPPAAP---GRPAFVAATPARRDEALRGDRDGLLPCTPRPRRPAAGVAGPT 123
QY 135 AKURLALWTPP-----PTPRGSH-----SPRPREAPL-----161
DB 124 AAGRGI---PPAGTDPAPRPRGRDHRHPQAPRHHPCRLRQPRGSPRVPAQLP 180
QY 162 ---RAIHPAGPSKTEPCPADSOK-----LSMGGHLSRPLKRLGSLHSLTHLNVPSG 211
DB 181 PRRADLAQSPAQGGGGAAGRRHAGNFCAGHGAIASRPAQ-----SG 227
QY 212 HPATSAPH-----TNG---PDLRPSTSGVTFR-----SPLV-TSRARSVSISV 251
DB 228 APAGHLPRPRAARGRRRTGGGQQPQRRRAVRLGLIKRRAGLPAPVAGTSTA-----280
QY 252 PSTPRCGATCKPXP 267
DB 281 --TP-PCRPTAFLPP 293

RESULT 14
US-09-252-991A-17829
; Sequence 17829, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17829
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17829

Query Match
Best Local Similarity 26.3%; Score 123.5; DB 4; Length 390;
Matches 75; Conservative 17; Mismatches 108; Indels 85; Gaps 14;

QY 22 RGYRVKARTSYVDETLFGSPAGTRPTPPDFDPWVEKANTRGVGKKEASKALGAKGSC 81
DB 80 RGDRLADLT-----GAQPTPL---PPGHVRLRLGGAAGAAAPATGAGARGE 124
```

```
QY 82 TTPSRGSTPTLTPRKKNKYRPISHTPSYCYDESIFGSRSEG---ASFGAPRMAKGDAAKLR 138
DB 125 LKDSFGTAPGNRPSGRRR-----TAGLRAGGIPPPASGGRRAPVRGA-FR 169
QY 139 ALLWTPPTTPRGSHSRPRE-APLRAIHPAG-----PSKTEPGPA---ADSQKLSMG 186
DB 170 LRLSAQPVRRPRRPPVRLPGGPPRARLATQPGKRGSRQRPGKTRPAPAGGADPAALHRG 229
QY 187 ---CLHSSRLPKRLGSLHSLTH-LNVPSG-----GHP-----213
DB 230 PQAAGRHH--RP---GTDHRRTHPRKVAATRTGTALRAAAGDFALRLCADLARRRRRSGR 284
QY 214 --ATSAPHTNGPDLRPTSGVTFRSPLVTISRARSVSISVSTPR 256
DB 285 RLAACTDPGGAEDLKKSQEGFSLRSISPTRSPSISLADSLRPR 329

RESULT 15
US-08-317-310A-64
; Sequence 64, Application US/08317310A
; Patent No. 5858701
; GENERAL INFORMATION:
; APPLICANT: WHITE, Morris F.
; APPLICANT: SUN, Xiao Jian
; APPLICANT: PIERCE, Jacalyn H.
; TITLE OF INVENTION: THE IRS FAMILY OF GENES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,310A
; FILING DATE: 03-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: JDP-022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1321 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-317-310A-64

Query Match
Best Local Similarity 26.3%; Score 123.5; DB 2; Length 1321;
Matches 72; Conservative 30; Mismatches 135; Indels 37; Gaps 11;

QY 4 PVELAVSGMOTLGLQHRCGGVYKARTSYVDETLFGSPAGTRPTPPDFDPWVEKANRT 63
DB 882 PTLISLEGLQTLPSMQEYPLPTEPKSPGYINID-FGE-AGTSLSP--APPLASAS 937
QY 64 RGVGKEASKA--LGAKGSCETTPSRGSTPTLTPRKKNKYRPISHTPSYCYDESIFGSR---118
DB 938 SLLSASSPASSIGSGTPTGTSRQSRQSLSDYMLDFSSPKSPKSPSTRSGDTVSGMDGL 997
QY 119 -SEGASFGAPRMAKGDAAKLRALLWTPPTPRGSHSRPREAPLRAIHPAGS---KTEP 174
DB 998 LSPERS--SP-----YPLPFPSTSPSSLQQPL-----PPAFCDLYRLPP 1036
```

QY 175 GPAADSOGLSMGGLHSSRPLKRGLSHSLTHLNVSTGHPATSAHTNGPQDLRPSTSGVT 234  
Db 1037 ASAATSOQFTAGSSMSSEPCDNGDYTEMAFGVAATPPQPIVAPPKEGARVASP-TSGLK 1095  
QY 235 FRSP--VTSRAKSVSISVPSTPRRGATQKXP 266  
Db 1096 RLSMDQVSGVEAFLOVSQPPDFHRCVKIRADP 1129

Search completed: August 10, 2004, 12:14:20  
Job time : 20 secs

This Page Blank (uspto)



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 12:10:47 ; Search time 46 Seconds  
(without alignments)  
1834.363 Million cell updates/sec

Title: US-10-031-589-4  
Perfect score: 1437  
Sequence: 1 MKTPVELAVSGMQTGLQHR.....SVPTFRGGATQKPKPWK 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1216	84.6	229	16	US-10-452-858C-79
2	239	16.6	467	12	Sequence 79, Appl
3	218	15.2	65	11	Sequence 1711, Ap
4	150	10.4	668	16	Sequence 8426, Ap
5	139	9.7	429	16	Sequence 134726, Sequence 134726,
6	136.5	9.5	4019	9	US-10-437-963-114012
7	136.5	9.5	4019	9	US-09-738-973-425
8	136.5	9.5	4019	9	US-09-854-133-425
9	135	9.4	442	12	US-10-144-649A-425
10	134.5	9.4	503	13	US-10-425-114-40452
11	134.5	9.4	19652	15	US-10-084-846A-7
12	134	9.3	625	16	US-10-437-963-143626
13	134	9.3	937	16	US-10-437-963-156445
14	133	9.3	380	16	US-10-437-963-200034
15	133	9.3	514	16	US-10-437-963-111701, Sequence 111701,

16	132.5	9.2	498	12	US-10-365-742-20	Sequence 20, Appl
17	132.5	9.2	497	16	US-10-437-963-125004	Sequence 125004,
18	132	9.2	621	16	US-10-437-963-177697	Sequence 177697,
19	131.5	9.2	277	16	US-10-437-963-177998	Sequence 177998,
20	130.5	9.1	417	16	US-10-437-963-143835	Sequence 143835,
21	130.5	9.1	878	16	US-10-722-050-2	Sequence 2, Appli
22	130	9.0	307	16	US-10-437-963-181279	Sequence 181279,
23	130	9.0	391	16	US-10-437-963-181279	Sequence 187780,
24	129.5	9.0	791	12	US-10-170-385-57	Sequence 57, Appl
25	129.5	9.0	863	15	US-10-359-012-2	Sequence 2, Appli
26	129.5	9.0	863	15	US-10-084-846A-5	Sequence 14, Appl
27	129.5	9.0	19723	15	US-10-078-547-24	Sequence 5, Appli
28	128	8.9	507	13	US-10-437-963-176449	Sequence 195781,
29	128	8.9	514	16	US-10-437-963-195781	Sequence 10907, A
30	128	8.9	797	14	US-10-156-761-10907	Sequence 160033,
31	127.5	8.9	388	16	US-10-437-963-160033	Sequence 150342,
32	127.5	8.9	713	16	US-10-437-963-150342	Sequence 120, App
33	127.5	8.9	1870	16	US-10-408-765A-120	Sequence 145797,
34	127	8.8	238	16	US-10-437-963-145797	Sequence 176449,
35	127	8.8	481	16	US-10-437-963-176449	Sequence 4, Appli
36	126.5	8.8	1321	16	US-10-122-805-4	Sequence 143, App
37	126	8.8	696	14	US-10-121-988-143	Sequence 143, App
38	125.5	8.7	699	14	US-10-300-562-143	Sequence 143, App
39	125.5	8.7	699	14	US-10-237-551-143	Sequence 254, App
40	125.5	8.7	699	14	US-10-237-551-254	Sequence 166325,
41	125.5	8.7	1103	16	US-10-437-963-166325	Sequence 136216,
42	125.5	8.7	469	16	US-10-437-963-136216	Sequence 116009,
43	125	8.7	365	16	US-10-437-963-116009	Sequence 137918,
44	124.5	8.7	365	16	US-10-437-963-137918	
45	124.5	8.7	556	16	US-10-437-963-137918	

ALIGNMENTS

RESULT 1

US-10-452-858C-79  
; Sequence 79, Application US/10452858C  
; Publication No. US20040086945A1  
; GENERAL INFORMATION:  
; APPLICANT: Sreekrishna, Kotikanyadanam  
; APPLICANT: Gerwe, Gina S  
; APPLICANT: Toerner, Daniel R.  
; TITLE OF INVENTION: HAIRLESS PROTEIN-INTERACTING PARTNER COMPLEXES AND METHODS THERE  
; FILE REFERENCE: 8956P  
; CURRENT APPLICATION NUMBER: US/10/452,858C  
; CURRENT FILING DATE: 2003-06-02  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 79  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (217)..(217)  
; OTHER INFORMATION: The 'Xaa' at location 217 stands for Lys, Arg, Thr, or Met.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (219)..(219)  
; OTHER INFORMATION: The 'Xaa' at location 219 stands for Tyr, Cys, Ser, or Phe.  
; NAME/KEY: misc feature  
; LOCATION: (225)..(225)  
; OTHER INFORMATION: The 'Xaa' at location 225 stands for Leu, or Phe.  
US-10-452-858C-79

Query Match 84.6%; Score 1216; DB 16; Length 229;  
Best Local Similarity 98.7%; Pred. No. 2e-84;  
Matches 226; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 11 GMDTGLQHRGGRVYKARTSYDVFSGSAGTRTPDPDPDPWKEKANTRGVKEA 70  
|||||

```
Db 1 GMOTLGLQHRRCRGYVKAARTSYVDLTFGSPAGTRTPDPDPFWVEKANRTGKGEA 60
Qy 71 SKALGAKGSCETTPSRGSTPTLTTRKKNKYRPISTHTPSYCYDESLFGSRSEGSFGAPRMA 130
Db 61 SKALGAKGSCETTPSRGSTPTLTTRKKNKYRPISTHTPSYCYDESLFGSRSEGSFGAPRMA 120
Qy 131 KGDAAKRALWTPPTPRGSHSPRPREALRAIHAPGPKTEFGPAADSKQLSMGGLHS 190
Db 121 KGDAAKRALWTPPTPRGSHSPRPREALRAIHAPGPKTEFGPAADSKQLSMGGLHS 180
Qy 191 SRPLKRGLSHLTHLVNSTCHPATSAHNTNGQDLRPSGTGVTFRSPL 239
Db 181 SRPLKRGLSHLTHLVNSTCHPATSAHNTNGQDLRPSGTGVTFRSPL 229

RESULT 2
US-10-276-774-1711
; Sequence 1711, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Tang, Y, Tom et al
; FILE OF INVENTION: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR FILING DATE: 2000-04-27
; PRIOR FILING DATE: 2000-04-27
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1711
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (1)...(467)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-1711

Query Match 16.6%; Score 239; DB 12; Length 467;
Best Local Similarity 29.4%; Pred. No. 5.6e-10;
Matches 94; Conservative 30; Mismatches 94; Indels 102; Gaps 17;

Qy 5 VELAVSGMOTLGLQHR--CRGGYR-----VKARTSYVDLTFGSPAGTRTPDPDPFW 56
Db 176 MRCLAAQAQAGLPHRTSIRPGWRRLTEPEAWARR---HRRFWGQRCVAVRPPQGAAPP 232
Qy 57 VEKANRTGKGEASKALGAKGSCETTPSRGSTPTLTTRKKNKYRPISTHTPSYC---DBS 113
Db 233 SHQGRRT-----NTDPS--ATPLRT-----VNGRCLAPDLK 261
Qy 114 LFGSRSEGSFGAPRMAKGAAXLRALLWTPPTPRGSHSPRPREALRAIHAPGPKTE 173
Db 262 APASGPRGWRMRGMPQS-----GALLWTPPTPRGSHSPRPREALRAIHAPGPKSR 314
Qy 174 PGPAADSKQLSMGGLHSRPLKKG-----LHSL-----THLVNSTCHPAT 216
Db 315 AGASGRLEPIYGVNLTFTPEAGTFLIPSTXMSPALVIQPPVPTQGLSGLRQ 374
Qy 217 APTNGQDLRPSGTGVT-----RSLVTSRARSISVSPTRP----- 257
Db 375 XP--SGAPFWL--PGLAQIAQCHLPHDEVGPPRNQSLGND---TLSSGLPWGRQVWPL 429
Qy 258 ---GG--ATQKP-----KPPW 268
Db 430 ARVGGHSPREPQVLKKPLW 449

RESULT 3
US-09-864-408A-8426
```

```
; Sequence 8426, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; FILE OF INVENTION: No. US20040009474A1el Human Polynucleotides and Polypeptides Enco
; TITLE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8426
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-8426

Query Match 15.2%; Score 218; DB 11; Length 65;
Best Local Similarity 71.4%; Pred. No. 2.4e-09;
Matches 40; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Qy 12 MDTLGLQHRRCRGYVKAARTSYVDLTFGSPAGTRTPDPDPFWVEKANRTGKVG 67
Db 1 MKAHLQHRSTSYRVKARASYVDLTFGSPARTPAQDPDFPFWQNCNRSRGVG 56

RESULT 4
US-10-437-963-134726
; Sequence 134726, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 134726
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_36471C.1.pep
US-10-437-963-134726

Query Match 10.4%; Score 150; DB 16; Length 668;
Best Local Similarity 26.7%; Pred. No. 0.0047;
Matches 72; Conservative 21; Mismatches 113; Indels 64; Gaps 13;

Qy 23 GGVYVKAARTSYVDLTFGSPAGTRTPDPDPDP--PWVEKANRTGKGEASKALGAKGSC 81
Db 80 GGVNPPSPSITGPTTPGGGGYVTPTPSDTPPSPSDTSPSTPGGCSSTP-----PCD 134
Qy 82 TTPSRGS--TTLTPRKNKYRPISTHTPSYCYDESLFGSRSEGSFGAPRMAKGAAXLRAL 140
Db 135 APPSPSDTSPSTPGGGGYSP---TFS-----DTPPSPSD----- 168
Qy 141 LWTPTPTP--RGSHSPRPREAP-----LRAIHAG-----PSKTEPGPAADSKQLSM 185
Db 169 --TSPTTPGGGGYVTPTPSDAPPSPSDTSPSTPGGGGYVTPTPSDAPPSPSDTSPPTP 226
Qy 186 GGLHSSRPPLKGLSHSLTHLVNSTCHPATSAHNTNGQDLRPS-----GVTRF 236
```

```
Db 227 GGGGGYTPPTPPSPSSGSPPTT--PGGGGYTPTSDTTPSPSSGSRRTTPGGCSTP 284
QY 237 SLVTSRARSVSISVPSPTRRGATQKPKP 266
Db 285 TPGTTPAPSSGTS-PTTP--GGSYPPTP 311

RESULT 5
US-10-437-963-114012
; Sequence 114012, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 114012
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(429)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17745C.1.pep
US-10-437-963-114012

Query Match 9.7%; Score 139; DB 16; Length 429;
Best Local Similarity 24.5%; Pred. No. 0.02;
Matches 83; Conservative 36; Mismatches 100; Indels 120; Gaps 19;

QY 21 CRGGYR---VKARTSYVDTEFGS-----BAG-----TRPTPD-----FDPWPWEKAN 61
Db 5 CTGVRWFVRVRRRPRYTHRQLLTQSRPHPPPAAPLVPSQPPPPDAPLAPIRPPSIGCAS 64
QY 62 RTRGVGXEASKALCAGKS-----CETTPSRG-----STPTLTTRKXNKY 100
Db 65 RAKPASISAAASAEFAFAAIGCTSPSRPIPPENDAMPARCNELPRPBLAVRGRHAS 124
QY 101 RPISH-----TPSYCDESIFGSRSEGASFGAPRMKDAKRLALLTPPTPRGSH 152
Db 125 NDRSHHRRSRRCCQMPSHLAILTS-SAAATXRSRPS-QAVRCRSL---QPPAP--- 175
QY 153 SPRPREAPLRAIHGAPSKTEPGPAADSQKLSMGGLH-----SSRPLKRGLSH- 200
Db 176 -----AVAAELHVAGDSILPPP-----HVAGLHDAC TAGLAHAEARPPRCILDH 223
QY 201 -----SUTHLNVPSTGHPATSA--PHNPGQDLRPTSGVTFRSPL--VTSRAR----- 245
Db 224 GRCLASQSQSQMPQPPQPVTAAGVSHNRQPDL---TRGVVASSCIAGVTGSRXPPPP 280
QY 246 -----SVSISVPSPTRRGATQKPKP 267
Db 281 HLSSGRASVLGTGSGSGSVXKXVATAPGR-----RRPAPP 315

RESULT 6
US-09-738-973-425
; Sequence 425, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
```

```
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather.
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-425

Query Match 9.5%; Score 136.5; DB 9; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.38;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDTEFLFGSPAGTR-----PTPDPFDPWPWEK-----ANTRGVGK 68
Db 988 SRPLQMNETTANRPSPVRDLGSSSTTNNDPYAKPDTPRPVMTDQFPKSLGSLRSPVSE 1047
QY 69 EASKALGAKGSC-----ETTPSRGSTTLTPRKNK-----YRPISTPS 108
Db 1048 QTAKGPIAAGTSDHFTKPSRADVFQRIQIPDSVARLLTAPLDSEGGPKTQMOPPS 1107
QY 109 YCDESIFGSRSEGASFGAPRMKDAKRLALLTPPTPRGSH--SPRPREAFLRAIHP 166
Db 1108 SQDP--YGSVSQ-----ASRRLSVDPYERPAL--TPRIDNFHSHNQNDPYSQPLTPHP 1158
QY 167 -----AGPSK--TEPG-----PAADSQKLSMGGLHS-----SRPL 194
Db 1159 AVNESFAHPSRAFSQPGTISRPTSQDPYSQPPGTPRPFVVDYSOSSGTPARENTPYSQP- 1217
QY 195 KEGLSHLTHLNVSTGHPATSAHPNPGQDLRPTSGVTFRSPLVTSRARSVSISVPSST 254
Db 1218 -----POTPRPTTVDPYSQQPQTPRPRSTQTDLFVTPVTHQHSDDPYAHPPTG 1264
QY 255 PRGGATQKPKP 267
Db 1265 PRGGSVPSYQPP 1277

RESULT 7
US-09-854-133-425
; Sequence 425, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
; TYPE: PRT
```

```

; ORGANISM: Homo sapiens
US-09-854-133-425

Query Match          9.5%; Score 136.5; DB 9; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.38;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPAGTR-----PTPPDFDPWVEK-----ANRTGVGK 68
DQ 988 SRPLQWNETTANRPSPVRDLCSSTTNNNDPYAKPPTPRFVMTDQPKSLGSRSPVSE 1047
QY 69 EASKALGAKGSC-----ETTPSRGSTTTLTPRKNK-----YRPISHTPS 108
DQ 1048 QTAKGPIAAGTSDHFTKPSRADVFQRIQIPDSVARELLTPALDSGPGFKTPMQPPPS 1107
QY 109 YCDSELFGRSEGFAGPRVAKGDAKRLALLWTPPTPRGSH--SPRPREAPLRAIHP 166
DQ 1108 SQDP--YGSVSQ-----ASRRLSVDPYERPAL--TPRPIDNFHNSQNDPYQPPLTTPH 1158
QY 167 -----AGPSK--TEPG-----PAADSQKLSMGGLHS-----SRPL 194
DQ 1159 AVNESFAHPSRAFSQGTISRPTSQDPYSQPPGTPRPVVDYSQSSGTSARSTNTDPYSQP- 1217
QY 195 KRGLSHSLTHLNVSTGHPTATSAHTNGPQDLRPSTSGVTFRSPLVTSRARSVSISVPST 254
DQ 1218 -----PGTTPRTTVDYPSQPPQTPRSTQTDLFVFTVNTQRHSDPYAHPPGT 1264

QY 255 PRRGATQKPKPP 267
DQ 1265 PRPGISVFPYQPP 1277

RESULT 9
US-10-425-114-40452
; Sequence 40452, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 2003-04-28
; CURRENT FILING DATE: 2003-04-28
; SEQ ID NO 40452
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB143-063-E8_FLI.pep
US-10-425-114-40452

Query Match          9.4%; Score 135; DB 12; Length 442;
Best Local Similarity 28.5%; Pred. No. 0.041;
Matches 69; Conservative 28; Mismatches 101; Indels 44; Gaps 12;

QY 37 TLFGSPAGTRTPPPDFPPWVEKANRTRGVGKEASKALGAKGSCETTPSGSTPTLTPR- 95
DQ 55 TTRCPRGTSPSPPTSPSTSSAPRS-----PRPATTSPPTGARCPRPPRW 102
QY 96 KKNYRISHTPSYCDSELFGRS-EGASFGAPEMAKDAKRLALLWTPPTPRGSHSP 154
DQ 103 RSRRRPPSSAS-----ARSGAAAATAAPR-----APSPATATSTCTSPR---SP 145
QY 155 RPREAPLRAIHPAGPSKTEPG---PAADSQKLSMGGLHSRPLKRLGSLHSLHNVSTG 211
DQ 146--TRTPTTRASRASPAAASPCSPSPRPPSPATAGTSTTRSRSLR--CSTACPGRCPTSG 203
QY 212 HPATSAPH--TNGPQDLRPST--SGVTFRSP-----LVTSPARSVSISVSTPREGGATQX 263
DQ 204 ACATWSPSTCTPGRAPTPPPGTRGSSAPSPASSSTASTSSRSVGAATCSARRRRRRR 263
QY 264 PK 265
DQ 264 PR 265

RESULT 10
US-10-078-547-2
; Sequence 2, Application US/10078547
; Publication No. US20020199211A1
; GENERAL INFORMATION:
; APPLICANT: Narayanasamy Ramesh
; APPLICANT: Miguel A. de la Fuente
; APPLICANT: Ines M. Anton
; APPLICANT: Ralf S. Geha
; TITLE OF INVENTION: WIP, A WASP-Associated Protein
; FILE REFERENCE: 1242.1022-005
; CURRENT APPLICATION NUMBER: US/10/078,547

```

```

; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/599,287
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/068,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 503
; TYPE: PRP
; ORGANISM: Human
US-10-078-547-2

Query Match          9.4%; Score 134.5; DB 13; Length 503;
Best Local Similarity 25.8%; Pred. No. 0.052;
Matches 63; Conservative 23; Mismatches 107; Indels 51; Gaps 11;

QY      42 PACGTR-TTPDFDPPWVEKANRGTGVCKEASKALGAKGCETTPSRGSTPT--LTPRKGN 98
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
Ddb     207 PGGPQPSPSGTPTPPFPFNGNRGTALGGGSIQSPLSSSPPFNRPPLPTPSRALDDKPPP 266
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      99 KYRISHTPSYCDSSLFGSRSEGAFCAPRMAGDAAKLRALLWTPTTPRGSHSPRRP- 157
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
Ddb     267 PPPVGVNPEISHRAV-----PPPPQNKKPVVPST 297
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      158 ---RAPURA-IHPAGPKTBPFGPAADSQKUSMGLHSSRLPKKLGHSLTHLNVPSTGH- 212
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
Ddb     298 PRPSAPHPRHLRPPPPSRPGPPPLPPS---SSGNDETPLRPOKNLSLSSSTPLPSPGSR 354
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      213 -PATSAPTHNGPDQLR--PSTSGVTFRPLVTSRASVSISVPSTPR---RGGATQK--- 263
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
Ddb     355 GPLPPPPGERPPVPURDPGRSG-PLPPPPVSRNGSTSEALPATQLPSRSGVDSPRSG 413
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      264 PKXP 267
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db       414 PRPP 417
         |||:|||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:
;
RESULT 11
US-10-084-846A-7
; Sequence 7, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 7
; LENGTH: 19652
; TYPE: PRP
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.
; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.
US-10-084-846A-7

Query Match          9.4%; Score 134.5; DB 15; Length 19652;
Best Local Similarity 25.7%; Pred. No. 3.3;
Matches 80; Conservative 29; Mismatches 103; Indels 99; Gaps 16;

```

QY 172 -----TERCPAADSOKLSNGGLHSSRRPLKRGSLHSLTHLNV-----STGHPATSA 217  
Db 386 FGRPACPIILLSPSPASISLAASVA---CPAPSPSTLPRTRDPSPPXSTAGAPPSTA 442  
QY 218 --PHNGQDLRPSSTGVTFPSPLVTSRAR-----SVSISV-----PST-PR 256  
Db 443 SPPPLISASLRRSPSPSPPLVTTISRQSTSPSPPPPSISLAASVACAPSPSTLPR 502  
QY 257 RGGATQKPKPP 267  
Db 503 RTRPSPSPSP 513

## RESULT 13

US-10-437-963-156445  
; Sequence 156445, Application US/10437963  
; Publication No. US20040123343A1

## GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 156445

; LENGTH: 937

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(937)

; OTHER INFORMATION: unsure at all Xaa locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_56111C.1.pep

US-10-437-963-156445

Query Match 9.3%; Score 134; DB 16; Length 937;  
Best Local Similarity 26.2%; Pred. No. 0.11;  
Matches 68; Conservative 21; Mismatches 83; Indels 88; Gaps 14;

QY 40 GSFAG-TRPTTPDPDPFWVEKANTRGVGKEASKALGAKGSCETTPSRG-----88  
Db 702 GPPPGSRPTQP---PEPLSRVTR-----VDAARPTASRASPPHLSPPS 747

QY 89 -----TPTLTRKKNKYPISHTPSYCDSESLF-GSRSEGASFGAPRMAKGDAAKL 137  
Db 748 SLSRAAGRLQPLSPF-PSAARPCRLRFRRRRLFPAGAAGNSWGNHRRRLVAGTKVAL 806

QY 138 RALLWTPPTTPRGSHGPRPRE-APLR---AIHPAGPS-----KTEPGAADSOQLSMG--G 187  
Db 807 PPLYCQLPPLPRSGAGRLRLPLRRRVALHGLSVLPERRP-----RRLSLGLVG 860

QY 188 LHSRLPKRGLSHLTHLNVSTGCHATSAPHTNGPDQLRPSSTGVTFPSPLVTSRARSV 247  
Db 861 THSGSPFR-----ASAPADVAVSTSGIA-----AGKV 888

QY 248 SISV---PSTPRRGATQKP 264

Db 889 VLGIASPPSKFLRGPSSPP 908

## RESULT 14

US-10-437-963-200034

; Sequence 200034, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 200034  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(380)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_95543C.1.pep  
US-10-437-963-200034

Query Match 9.3%; Score 133; DB 16; Length 380;  
Best Local Similarity 24.2%; Pred. No. 0.049;  
Matches 62; Conservative 16; Mismatches 114; Indels 64; Gaps 9;

QY 41 SPAGTRPTTPDPDPFWVEKANTRG-----VGKEASKALGAKGSCETTPSRG 87

Db 86 SAAAAAPPPTAAPPHRRRRCRCGNPPPLPPSPSSAASAPSRARSPPAATPCPARP 145

QY 88 STPTLTRKKNKYPISHTPSYCDSESLFSGRSEGASFGAPR---MAKGAALKRALLWT 143

Db 146 RCGSATP--PAPWRXPSPSPAPAPAGCAARTPPPAYVHRLSMTIPPPPRHHAAL 203

QY 144 PPPTPR---GSHSPRPREAPLRAIH-----PAGPSKTEPGAADSOQLSMGLHSSRP 193

Db 204 PPAPPRPNRTTRPRPARRRHRLRIPPPPLPAPPSPAPPGP-----246

QY 194 LKRGSLHSLTHLNVSTGCHATSAPHTNGP--ODLRPSTGVTFPSPLVTSRARSVSISV 251

Db 247 -----RHPPFPSP-HPPXPQPRPRPPAGRNHLHSPPPPNPPLPPSR-----GP 290

QY 252 PSTPRRGATQKPKPP 267

Db 291 PPSPSRSPATDLPPAP 306

## RESULT 15

US-10-437-963-111701

; Sequence 111701, Application US/10437963

; Publication No. US20040123343A1

## GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 111701

; LENGTH: 514

```
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(514)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_15655C.1.pap
; US-10-437-963-111701

Query Match          9.3%; Score 133; DB 16; Length 514;
Best Local Similarity 26.6%; Pred. No. 0.069;
Matches 64; Conservative 20; Mismatches 117; Indels 40; Gaps 9;

QY 42 PAGTRTPDPPDPFWVEKANRTRGVGKEAS-----KALGAKGSCETTPSRGSTPTLTP 94
Db 239 PARTLPAPPXADPPGRGGXPRRRATGTRAAPPFACFTRRSFAGRGOWSPRPPPGSGSPFAP 298
QY 95 RKCKKYRPI-SHTPSYCDLSFGSRSEGSFGAPRMAKGDAAKLRALLWTPPTPRGSHSP 154
Db 299 -----PPPPPPRGAPRAAPAASAAPASAAGRAPRAPRPPRPPPPPSARASSSP 352
QY 155 RPREAFLRA----IHPAGPSKTEPPGPAADSQKLSMGLHSSRPLKKGLSHSLTHLNVST 210
Db 353 RHSATPAPAHXPRPHSPGCGRPPTTRAPSS-----PWPPPTGAAP-----RGPT 398
QY 211 GHFATGAPHTNG-PQDLR---PSTSGVTFPSPLVTSRAESVTSVSTPRGGATOKPKP 266
Db 399 PAPSSSWPPPPAGTFAALRWAAAPASAPCTACS--AAGRARG---SSPPPPQHRHHHSPP 453
QY 267 P 267
Db 454 P 454
```

Search completed: August 10, 2004, 12:13:50  
Job time : 47 secs

This page Blank (uspto)



LTR gag/pol polypr  
toucan gene protei  
calcium channel pr  
hypothetical prote  
synapsin Ia - rat  
EBNA-LP protein -  
A4 protein - mous  
nascent polypeptid  
tpr homolog - fhu  
hypothetical prote  
hypothetical prote  
DNA-binding protei  
extensin-like cell  
proline-rich cell  
hydroxyproline-ric  
Bfuf1 protein - hu

30 115.5 8.0 1456 2 T01397  
31 115.5 8.0 2176 2 T13806  
32 115.5 8.0 2212 2 A41098  
33 115 8.0 380 2 D70516  
34 115 8.0 704 2 A30411  
35 114.5 8.0 308 2 S77938  
36 114.5 8.0 1211 2 T42230  
37 114.5 8.0 2187 2 T30826  
38 114 7.9 2346 2 T13829  
39 113.5 7.9 346 2 T46916  
40 113 7.9 1952 2 T48814  
41 112.5 7.8 2282 2 T42717  
42 112 7.8 214 2 T10737  
43 112 7.8 214 2 T09854  
44 112 7.8 620 2 S06733  
45 112 7.8 3149 1 Q08E8

ALIGNMENTS

RESULT 1

T28770  
hypothetical protein W03D2.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T28770  
R:Kohling, T.; Wohlmann, P.  
submitted to the EMBL Data Library, June 1997  
A:Description: The sequence of C. elegans cosmid W03D2.  
A:Reference number: Z20519  
A:Accession: T28770  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-539 <ROW>  
A:Cross-references: EMBL:AF000298; PIDN:AAC48255.1; GSPDB:GN00022; CBSP:W03D2.1  
A:Experimental source: strain Bristol N2; clone W03D2  
C:Genetics:  
A:Gene: CBSP:W03D2.1  
A:Map position: 4  
A:Introns: 40/3; 88/3; 115/3; 146/3; 173/3  
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Result No.	Score	Query Match	Length	DB ID	Description
1	144.5	10.1	539	2 T28770	hypothetical prote
2	136	9.5	2715	2 T13049	eyelid - fruit fly
3	132.5	9.2	580	2 T43481	probable mucin DKF
4	130.5	9.1	705	2 A35363	synapsin I splice
5	128	8.9	403	2 S52796	prpL2 protein - hu
6	127.5	8.9	1870	2 S37671	MHC class III hist
7	127.5	8.9	1872	2 S36152	MHC class III hist
8	127	8.8	302	2 T32711	hypothetical prote
9	126.5	8.8	1199	2 A40670	nuclear envelope p
10	125.5	8.7	706	2 B30411	synapsin Ia - bovi
11	125.5	8.7	699	2 C43674	US4 protein - huma
12	124	8.6	576	2 T36729	probable serine/th
13	123.5	8.6	724	2 T47149	hypothetical prote
14	123	8.6	886	2 S29605	glycoprotein 350/2
15	122.5	8.5	907	1 Q08E21	membrane antigen 9
16	122	8.5	1777	2 T34369	hypothetical prote
17	121	8.4	721	2 E70766	high molecular mas
18	121	8.4	1151	2 T18535	MHC class III hist
19	120.5	8.4	2142	2 B35098	NIMA-like protein
20	120	8.4	779	2 A37177	unconventional myo
21	120	8.4	3530	2 A9266	immediate-early pr
22	119	8.3	1446	1 A45344	serine-rich protei
23	118.5	8.2	1077	2 A44067	En/spm-like transp
24	116.5	8.1	306	2 T52305	unconventional myo
25	116.5	8.1	3511	2 A59295	hypothetical prote
26	116	8.1	635	2 F75477	hypothetical prote
27	115.5	8.0	846	2 T21700	hypothetical prote
28	115.5	8.0	883	2 A96662	hypothetical prote
29	115.5	8.0	1285	2 T14171	ataxin-2 - mouse

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: August 10, 2004, 12:08:42 ; Search time 17 seconds  
(without alignments)  
1522.089 Million cell updates/sec

Title: US-10-031-589-4  
Perfect score: 1437  
Sequence: 1 MKTPVELAVSGMOTLGLQHR.....SVPTPRRGATQKPKPKWK 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

T13049  
 eyelid - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C:Accession: T13049  
 R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.  
 submitted to the EMBL Data Library, March 1998  
 A:Reference number: Z17592  
 A:Accession: T13049  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2715 <TRE>  
 A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1  
 C:Genetics:  
 A:Gene: eld  
 A:Cross-references: FlyBase:FBgn0003013  
 C:Function:  
 A:Description: could act as a transcription factor antagonistic to the wg pathway  
 C:Keywords: DNA binding

Query Match 9.5%; Score 136; DB 2; Length 2715;  
 Best Local Similarity 22.5%; Pred. No. 0.41; Indels 64; Gaps 11;  
 Matches 62; Conservative 30; Mismatches 119; Indels 64; Gaps 11;

QY 23 GGYRV-----KARTSYVDETLFGSPAGTPTPPDF-----DPPWVEKANRTRGVG 67  
 DB 632 GGYKMGPGQSPGAGQYPPQPPQYPPGNYPPRPQYPPGAYATGPPPPPTSQA-----GAG 687  
 QY 68 KEASKALGAKGSCETTTSRSGSTPLTPRKKNKYRPISHTPSYCDLSFGSRSGASFGAP 127  
 DB 698 GANSMPSGAQAG--GYPCRG-----MENHTGQYPPYQWVPSPQQTVPFGAPGGMVGNH 740  
 QY 128 RMAKGDAKLRLALLWTP-----PPTPGSHSPRPREAPLRAIHPAGPKTEPGPAADS 180  
 DB 741 VQKGG-----TPPPVVGPPPPQSGSPRLNLYLKQHLQKGGVGGSTPPQGP 790  
 QY 181 QKLSNG--GLHSSRL-----KRLSHSLTHLVNFT-----GHPATSAPHT 220  
 DB 791 QGYNGPTGMHPGMPGPPHMGPPHGGTNNGGTSTPPQSQMLQGGQGGGSGGPES 850  
 QY 221 NGPDLPSTSGVTFSPVLTTSRARSVSISVPSTP 255  
 DB 851 GGPEHI--SQDNGISSSGTCAAGMHAVTSVVTGGP 884

RESULT 3  
 T43481  
 probable mucin DKFPz434C196.1 - human (fragment)  
 N:Alternate names: protein DKFPz434B0635.1  
 C:Species: Homo sapiens (man)  
 C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 20-Apr-2000  
 C:Accession: T43481; T34549; T17264  
 R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, December 1999  
 A:Reference number: Z22514  
 A:Accession: T43481  
 A:Molecule type: mRNA  
 A:Residues: 1-580 <AAA>  
 A:Cross-references: EMBL:AL133561; NID:g6599133; PIDN:CAB63715.1; PID:g6599134  
 A:Experimental source: adult testis; clone DKFPz434C196  
 R:Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, October 1999  
 A:Reference number: Z21540  
 A:Accession: T34549  
 A:Molecule type: mRNA  
 A:Residues: 262-580 <POU1>  
 A:Cross-references: EMBL:AL120269; NID:g6102864; PIDN:CAB59245.2; PID:g7018420  
 A:Experimental source: adult testis; clone DKFPz434B0635  
 R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, September 1999  
 A:Reference number: Z18723  
 A:Accession: T17264  
 A:Molecule type: mRNA

A:Residues: 262-580 <POU2>  
 A:Cross-references: EMBL:AL117481; NID:g5911958; PIDN:CAB55954.1; PID:g5911959  
 A:Experimental source: adult testis; clone DKFPz434B061  
 C:Genetics:  
 A:Note: DKFPz434C196.1; DKFPz434B0635.1; DKFPz434B061.1

Query Match 9.2%; Score 132.5; DB 2; Length 580;  
 Best Local Similarity 26.3%; Pred. No. 0.14;  
 Matches 77; Conservative 30; Mismatches 125; Indels 61; Gaps 17;

QY 3 TPVELAVSGMQLGQHRCRGYRVKATSYV----DETLFGSPAGTPT--PDPFDPW 56  
 DB 46 TFRASLTQWETAL-----LRTLPRASLMTTTRASLMTTPRASPTKPPRASPT 98  
 QY 57 VEKANRTGVGKEASKALGAKGSCETTTSRSGSTPLTPRKKNKYRPISHTPSYCDLSFG 116  
 DB 99 PSRASPTRLPRASPMGSPHRASPMRTTPRAS--PTGTSTASP----TGTF--SASFTG 151  
 QY 117 SRSEGASFGAPMA-----KGDAKL-----RALLWTPPTTPRGSHSPRPREAPLRAIHPA 167  
 DB 152 TPRASPTGTTPRAWATRSPTASLTTPSRASLTTRWPPRASPTTTP--PRES--RMGSHRA 209  
 QY 168 GPSKTEP--GPAADSQKLSMGLSHSRPLKRGSL--SHSLTHLVNVPSTGHPATSAHPINGP 223  
 DB 210 SPTTPPRASPTTRPPRAS-----PRTTPRESLRTSHRASPTRMPPRASPTTRPP-- 260  
 QY 224 QDLRSTSGVTFR--SPLVTSRARSVSISVPSTPRRGATQPK-----PPW 268  
 DB 261 ---RASPTGSPPRASPMTPPRAS-----PRTTPRASPTTTPSRASLTTRTPSW 304

RESULT 4  
 A35363  
 synapsin I splice form a - human  
 C:Species: Homo sapiens (man)  
 C>Date: 17-Aug-1990 #sequence\_revision 17-Aug-1990 #text\_change 01-Dec-2000  
 C:Accession: A35363; B35363; A35805  
 R:Suedhof, T.C.  
 J. Biol. Chem. 265, 7849-7852, 1990  
 A:Title: The structure of the human synapsin I gene and protein.  
 A:Reference number: A35363; MUID:90243651; PMID:2110562  
 A:Accession: A35363  
 A:Molecule type: DNA  
 A:Residues: 1-705 <SUE>  
 A:Cross-references: GB:M56371; GB:J05431  
 A:Accession: B35363  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-659, 'KASPAQAQP' <SU2>  
 A:Cross-references: GB:M58378; GB:J05431  
 R:Sauerwald, A.; Hoesche, C.; Oschwald, R.; Kilimann, M.W.  
 J. Biol. Chem. 265, 14932-14937, 1990  
 A:Title: The 5'-flanking region of the synapsin I gene. A G+C-rich, TATA- and CAAT-less,  
 A:Reference number: A35805; MUID:90368667; PMID:2118519  
 A:Accession: A35805  
 A:Molecule type: DNA  
 A:Residues: 1-125 <SAU>  
 A:Cross-references: GB:M55301; NID:g338655; PIDN:AAA60608.1; PID:g553654; GB:J05630  
 C:Genetics:  
 A:Gene: GDB:SVN1  
 A:Cross-references: GDB:119606; OMIM:313440  
 A:Map position: Xp11.23-Xp11.23  
 C:Keywords: actin binding; alternative splicing; phosphoprotein

Query Match 9.1%; Score 130.5; DB 2; Length 705;  
 Best Local Similarity 23.5%; Pred. No. 0.23;  
 Matches 63; Conservative 24; Mismatches 108; Indels 67; Gaps 9;

QY 40 GSPAGTPTPPDPPWVEKANTRG-----YKKEASKALGAKGSCETTTPSRGSTP 90  
 DB 454 GPPAQQRPPQGGPPQPGPQRPQRPQRPQRPQRPQRPQRPQRPQRPQRPQRPQRPQRP 511  
 QY 91 TLTPRKKNKYRPISHTPSYCDLSFGSRSGASFGAPRMAKGDAKLRLALLWTPPTTPR 149

Db 512 TSAPO-----OPASQAPPTQGGROSRPVAGFGAPPAPRPPASPSPORQAGPQATQ 566  
Qy 150 ----GSHSPRPREAPLRAIHPAGSKTEPGPAADSOQLSGVGLHSSRPLKRLGSLHSLTHL 205  
Db 567 TSVSGPAPPKASCAPPGGQORQGPQKPGPAGTPTQASQAG-----608  
Qy 206 NVSTGHPATSAHTNGP-----ODLRPSTSGVTFRSP---LVTSPARSV 247  
Db 609 PVFTGTGPTTQQPRPSGPGAGAPKPOLAQKPSQDVPPATAAAGGPPHQLNKSOGLTN 668  
Qy 248 SISVPSTPRRGATQKPKPP 267  
Db 669 AFNLP-----EPAPP 678

## RESULT 5

prpL2 protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 19-May-1995 #sequence\_revision 10-Nov-1995 #text\_change 10-Sep-1997

C:Accession: S52796

R:Ruhlmann, A.; Kreideweiss, S.; Nordheim, A.

submitted to the EMBL Data Library, March 1995

A:Reference number: S52796

A:Accession: S52796

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-403 &lt;RUH&gt;

A:Cross-references: EMBL:X86019; NID:g762950; PID:g762951

Query Match 8.9%; Score 128; DB 2; Length 403;  
Best Local Similarity 25.3%; Pred. No. 0.19;  
Matches 62; Conservative 23; Mismatches 108; Indels 52; Gaps 11;

Qy 42 PAGTR-PTPDDPDPVWEKANTRGVCKEASKALCAKGCETTPSRGSTPT--LTPRKN 98

Db 116 PGGPRQSPGPTPPFPFGNRTALGGGIRQSPSSSSPFSNRPPLPPTPSRALDOKPPP 175

Qy 99 KYRISHTPSYCDLSFGSRSEGFAGAPRMAGDAAKLRALLWTPPTTPRGSHSPRPR- 157

Db 176 PPPVGNRPISHRAV-----PPPPQNNKPPVPST 206

Qy 158 ---EAPLRA-IHPAGSKTEPGPAADSOQLSMGLHSSRPLKRLGSLHSLTHLNVSTGHP 213

Db 207 PRPSAPHRPLRPPPSRGPPLPPS---SSGNDETFLPQRLNLSLSSSTPLPSPGSR 263

Qy 214 A---TSAPHTNGPQDLR--PSTSGVTFRSPVTSRAESVTSVPSTPR---RGATQK-- 263

Db 264 GPLPPVPVSEPPPPVDRDPPGRSG-PLPPPPVSRNGSTRALPATPOLFSRSGVDSPRS 322

Qy 264 -PKPP 267

Db 323 GPRPP 327

## RESULT 6

S37671

MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human

C:Species: Homo sapiens (man)

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 15-Sep-2000

C:Accession: S37671

R:Bougueleret, L.

submitted to the EMBL Data Library, August 1992

A:Reference number: S37671

A:Accession: S37671

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1870 &lt;BOU&gt;

A:Cross-references: EMBL:Z15025; NID:g29374; PID:g29375

C:Genetics:

A:Map position: 6p21.3

A:Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 65

C:Superfamily: collagen alpha 1(IV) chain

Query Match 8.9%; Score 127.5; DB 2; Length 1870;  
Best Local Similarity 25.2%; Pred. No. 1;  
Matches 68; Conservative 17; Mismatches 112; Indels 73; Gaps 12;

Qy 33 YVDETILF-GSPAGTRPTPPDFDPWPV-----EKANTRGVGKE-----ASKALCAKGS 79

Db 728 YVDPRLLOGRP-----PLDFYPPGVHPGSLVPRERSDSRGLSSEPPFDRHAPAMLRERG 781

Qy 80 CETTPSR---GSTPTLTPRKNKYRPI-SHTFSPYCDLSFGSRSEGFAGAPRMAGDA 135

Db 782 PPVDFKLAWGDVFTATPAEP---RPLTSPLRQAADBDKGRSE-----823

Qy 136 KLRALLWTPPTPRGSHSPRPREAPLRAIHPAGSKTEPGPAADSOQLSMGG-----187

Db 824 -----TPVPV-----PYLASYPGFENGTPGPPISRFPLEEPGRPLPWP 867

Qy 188 -----LHSSRPLKRLGSLHSLTHLNVSTGHPATSAHTNGPQDLRPTSGVTFRSP 240

Db 868 GSDEVAKIQTTPPKKEPKKEETAQLTGPEAGKLPASRSAGAGPPPPPRESETRWGPRP 927

Qy 241 TSARSVSISVP-STPRRGATQKPKPPWK 269

Db 928 GSSRGIPPEEGAPRRRAGPIKKPPPTK 957

## RESULT 7

S36152

MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human

C:Species: Homo sapiens (man)

C:Date: 06-Jun-1995 #sequence\_revision 17-Nov-1995 #text\_change 15-Sep-2000

C:Accession: S36152

R:Irish, F.J.M.; Bougueleret, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Jurk

Nature Genet. 3, 137-145, 1993

A:Title: Dense Alu clustering and a potential new member of the NkappaB family within

A:Reference number: S36152; MUID:93272029; PMID:8499947

A:Accession: S36152

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1872 &lt;IRI&gt;

A:Cross-references: EMBL:Z15025

A:Note: in the authors' translation residues 32-34 are shown after residue 4 and, conse

A:Note: the authors translated the codon AAT for residue 1000 as His

C:Genetics:

A:Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 6

C:Superfamily: collagen alpha 1(IV) chain

Query Match 8.9%; Score 127.5; DB 2; Length 1872;  
Best Local Similarity 25.2%; Pred. No. 1;  
Matches 68; Conservative 17; Mismatches 112; Indels 73; Gaps 12;

Qy 33 YVDETILF-GSPAGTRPTPPDFDPWPV-----EKANTRGVGKE-----ASKALCAKGS 79

Db 729 YVDPRLLOGRP-----PLDFYPPGVHPGSLVPRERSDSRGLSSEPPFDRHAPAMLRERG 782

Qy 80 CETTPSR---GSTPTLTPRKNKYRPI-SHTFSPYCDLSFGSRSEGFAGAPRMAGDA 135

Db 783 PPVDPKLAWGDVFTATPAEP---RPLTSPLRQAADBDKGRSE-----824

Qy 136 KLRALLWTPPTPRGSHSPRPREAPLRAIHPAGSKTEPGPAADSOQLSMGG-----187

Db 825 -----TPVPV-----PYLASYPGFENGTPGPPISRFPLEEPGRPLPWP 868

Qy 188 -----LHSSRPLKRLGSLHSLTHLNVSTGHPATSAHTNGPQDLRPTSGVTFRSP 240

Db 869 GSDEVAKIQTTPPKKEPKKEETAQLTGPEAGKLPASRSAGAGPPPPPRESETRWGPRP 928

Qy 241 TSARSVSISVP-STPRRGATQKPKPPWK 269

Db 929 GSSRGIPPEEGAPRRRAGPIKKPPPTK 958

```

RESULT 8
T32711
hypothetical protein T22D1.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T32711
R:Geisel, C.; Bradshaw, H.; Hawkins, M.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid T22D1.
A:Reference number: Z21211
A:Accession: T32711
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-302 <GEI>
A:Cross-references: EMBL:AF039052; PIDN:AA94271.1; GSPDB:GN00022; CESP:T22D1.2
A:Experimental source: strain Bristol N2; clone T22D1
C:Genetics:
A:Gene: CESP:T22D1.2
A:Map position: 4
A:Introns: 99/1
C:Superfamily: proline-rich protein

Query Match      8.8%; Score 127; DB 2; Length 302;
Best Local Similarity 24.6%; Pred. No. 0.17;
Matches 65; Conservative 15; Mismatches 98; Indels 86; Gaps 11;

QY 44 GTRTPPPDFDP--WVEKANTR-----GVGKEASKALGAKG--SCETTPSRGSGTPTL 92
Db 72 GTTPPPPTGEPQDLSAEGNASRRPPPPKGTGTPPPPTGCEPQDLSGEGNASRRPP-- 129
QY 93 TPRKKNKVRPISTHTPSYCDLSFGRSEGASFGAPRMAKGAALKRALLTTPPTPRGSH 152
Db 130 -PPKGTGSPPPPT-----GEPQDLSGEGNASR-----RPPPPPKGTG 166
QY 153 SPRP-----REAPLRAIHAPGSKTEPGPAADSOGLSMGLHSSRPLK 195
Db 167 SPPPPPTGEPQDLSGEGNASRRPP-----PPKGTGTPPPPTGEPQDLSGEGNASRRP-- 220
QY 196 RGLSHLTHLNPVSTGHPATSAHTNGPQDLR-----FSTSGVTFRSPLVTSRAR 245
Db 221 -----PPPKGTGSGTPPTTGPQDLSGEGNASRRPPPPKGTGSGTPPPPTGEPQ 270
QY 246 SVSISVSTPRRGATQKPKPPWK 269
Db 271 DLS-----GEGNASRRPPPPK 287

RESULT 9
A40670
nuclear envelope protein POM 121 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 05-Nov-1999
C:Accession: A40670
R:Hallberg, E.; Wozniak, R.W.; Blobel, G.
J. Cell Biol. 122, 513-521, 1993
A:Title: An integral membrane protein of the pore membrane domain of the nuclear envelope
A:Reference number: A40670; MUID:93328754; PMID:8335683
A:Accession: A40670
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1199 <HAL>
A:Cross-references: GB:Z21513; NID:G396746; PIDN:CAA79725.1; PID:G396747
F:803-807,845-849,956-960,1010-1014,1047-1051,1076-1080/Region: pentapeptide motif (X-F-

Query Match      8.8%; Score 126.5; DB 2; Length 1199;
Best Local Similarity 26.1%; Pred. No. 0.74;
Matches 67; Conservative 26; Mismatches 131; Indels 33; Gaps 11;

QY 28 KARTSYVDETLF----GSPAGTRFTPPDFPPWVEKANTRGVGKEASKALGAKGSCETT 83
Db 362 RSTSSVSLTSTCTGIFSSRNA-----ITSSYSSTRGV-SQLWKRSGTSPSPSS 413
QY 84 P--SRGSTLTTPRKKNKVRPISTHTPSYCDLSFGRSEGASFGAPRMAKGAALKRALL 141

```

```

Db 414 PASSRSQTPE-RPAKTRREEEFCHOSSSSAPLVTKESPGKVKTPATGKQOS-----L 466
QY 142 WTPPTTPRGSHSPREAPL-----RAIHPAGSKTEPGPAADSOGLSMGLHSSRPLKRG 197
Db 467 WTSPTTP--GSSGQRKKIKLLPSRRGDQLTLPPPELGVSTITAEDLMERRASLQWENKV 525
QY 198 LSHSLTHLNPVSTG--HPATSAHTNGPQDLRPTSTSGVTF-----RSPVTSRARSVISI 252
Db 526 LEDXTDDASTPATDTSPTSPSTFTLTPTVGAASPASLPAPSSNPPLLESLLKMGQESP 585
QY 253 ST--PRRGATQKPKPP 267
Db 586 SSSPEPEAAATVAAPSP 602

RESULT 10
E30411
synapsin Ia - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 01-Dec-2000
C:Accession: E30411; F30411; A35758
R:Suedhof, T.C.; Czernik, A.J.; Kao, H.T.; Takei, K.; Johnston, P.A.; Horiuchi, A.; Kana
Science 245, 1474-1480, 1989
A:Title: Synapsins: mosaics of shared and individual domains in a family of synaptic ves
A:Reference number: A30411; MUID:89388265; PMID:2506642
A:Accession: E30411
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-661 <SUE>
A:Accession: F30411
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-661, 'KA', 664, 'PAQAQP', <SU2>
R:Hall, F.L.; Mitchell, J.P.; Vulliamt, P.R.
J. Biol. Chem. 265, 6944-6948, 1990
A:Title: Phosphorylation of synapsin I at a novel site by proline-directed protein kinas
A:Reference number: A35758; MUID:90216728; PMID:2108963
A:Accession: A35758
A:Molecule type: protein
A:Residues: 532-556 <HAL>
C:Keywords: actin binding; alternative splicing; phosphoprotein
F:551/Binding site: phosphate (Ser) (covalent) (by proline-directed kinase) #status expe

Query Match      8.8%; Score 126; DB 2; Length 706;
Best Local Similarity 22.9%; Pred. No. 0.46;
Matches 60; Conservative 23; Mismatches 87; Indels 92; Gaps 12;

QY 40 GSPAGTRFTPPDFPPWVEKANTRG-----VKEASKALGAKGSCETTPSRGSTP 90
Db 454 GPPAQARP-PPQGFPFGPGPQPGPPLQORPTPQOHLGLGPPAG-SPLPQRLPSP 511
QY 91 TLTPRKKNKVRPISTHTPSYCDLSFGRSEGASFGAPRMAKGAALKRALLTWP--PPTP 148
Db 512 TSVFQ-----QPASQ-----ATPMTQGGQRQSRPVAGGPGAPPAT 546
QY 149 RGSHPSPR-----EAPLRA--IHDA-----GPSKTEPGPAADSOGLSM 185
Db 547 RPPASPSFQAGPQATQTSVSGOAPPKASGVPPGGQQRQPGPQKPPGAGPTRQASQ 606
QY 186 GGLHSSRPLKRLSHLSLTHLNPVSTGHPATSAHTNGPQDLRPTSTSGVTSRARS 245
Db 607 AG-----PMRTGPTTQQRPSGPGPAGPTKPLAQKP----- 641
QY 246 SVSISVSTPRRGATQKPKPP 267
Db 642 SQDVPPATAAAGG-----PP 657

RESULT 11
C43674
US4 protein - human herpesvirus 2 (strain HG52)
C:Species: human herpesvirus 2

```

Db 406 APSAPSAPGPTRPAP-----HGTHSSEVFLAERPGAMSEFGSHLPPQPPTVT 456

Qy 210 -TCHPATs-----APH-----TNGPDQLRSTs-----GVTFSPIVTSR 243

Db 457 PTSDAAASDAAEAAQPPAPHPAFTGGPGLGPPDRAPGRSQHPAGHPGLIARS-LAPSP 515

Qy 244 ARGSVI-----SVPSTPRRGG-ATQKPKPP 267

Db 516 AREADVPTAAYTARNPRsAPPAQHRGARRRRPGPP 551

RESULT 13

T471149

hypothetical protein DKFZp547Pl03.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 01-Dec-2000

C:Accession: T471149; E56695

S:Bloeker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S. submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24376

A:Accession: T471149

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-724 <AAA>

A:Cross-references: EMBL:AL162059

A:Experimental source: fetal brain; clone DKFZp547Pl03

S:Stifani, S.; Blauweller, C.M.; Redhead, N.J.; Hill, R.E.; Artavanis-Tsakonas Nature Genet. 2, 119-127, 1992

A:Title: Human homologs of a Drosophila enhancer of split gene product define a

A:Reference number: A56695; MUID:93265135; PMID:1303260

A:Accession: E56695

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 277-418,'P',420-459,'A',461-724 <STI>

A:Cross-references: GB:M99439; NID:g307515; PIDN:AAA61195.1; PID:g307516

C:Genetics:

A:Gene: GDB:TLE4; ESG

A:Cross-references: GDB:228050

A:Note: DKFZp547Pl03.1

C:Superfamily: unassigned WD repeat proteins; WD repeat homology

C:Keywords: nucleus

F:480-513/Domain: WD repeat homology <WD1>

F:566-599/Domain: WD repeat homology <WD2>

F:648-681/Domain: WD repeat homology <WD3>

F:689-722/Domain: WD repeat homology <WD4>

Query Match 8.6%; Score 123.5; DB 2; Length 724;

Best Local Similarity 23.2%; Pred. No. 0.69;

Matches 67; Conservative 38; Mismatches 113; Indels 71; Gaps 13

Qy 3 TPVEL-AVSGMOTIGLHRCRCGGYEVKARTSYVDTLFGSPAGTRPTDPDPMPVVKAN 61

Db 71 TMAELNAIIGQQQQAQHLSHG-----HGLVPVLTTPHSGLOPPAIPIG 115

Qy 62 RTRGVGKEAKALGAKGSC-----ETTPSRGSTPTLTTPRKNKYRPIS 104

Db 116 SSAGL-LALSALGQSHLPKDKKHHNDHQDRDSDIKSSVSFSAFRAEKHR--- 171

Qy 105 HTPSYCDSELPGRSSEGAFCAPVAKGDAAKLALLWT---PPTPRGS-HSPR--- 155

Db 172 NSADYSSESK-KQATEKEIAARYDSGEKDDNLVDVSNEDSSSPRGSFAHSPPENGL 230

Qy 156 -----PREAPLRAIHAPGPKTEPGPAADSKLSMGGLHSSRPLKRGLSHSLTLNVPS 209

Db 231 DKTRLLKKDAP---LSPASIASSSSTPSSKSKELSLN-----EKSTTPVSKSN 275

Qy 210 TCHPATGA--PHTNGPDQLRSTSGVTFRSPVTSRAKSVSISVP-STP 255

Db 276 TPTPRDAPTSGSNSTPGLREVPGKPGVDPLASSLRTPMVPCPYPTP 324

RESULT 14

S29605





```

Db 240 TSAPO-----QVSAQAPLSAQGQSPVAGGFGAPPAARPPSPQFQAGPPQATRQ 294
QY 150 ----GSHSPRPREAPLRAHPAGSKTEPPGPAADSOQKLSMGGLHSSRPLKGLSHSLTHL 205
Db 295 TSVSGQAPFKASGAPSPGQKPPGPKPPGAGTRQASQAG----- 336
QY 206 NVPSTGHATPAGHTNGQDLRPSTGVTFRPLVTSRARSVSVSVPSTPRRGATOKPK 265
Db 337 PMPTGPTTQPP-----RPSGPGAGR-PAKPOLAQKPSQDVSPATAG----- 381
QY 266 PP 267
Db 382 PP 383

RESULT 2
MML3 HUMAN
ID MML3 HUMAN STANDARD; PRT; 4911 AA.
AC Q8NEZ4; Q8NCF6; Q9H9P4; Q9NR13; Q9P222; Q9UDR7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog (Histone-
DE lysine N-methyltransferase, H3 lysine-4 specific MLL3) (EC 2.1.1.43)
DE (Homologous to ALR protein).
GN MML3 OR HALR OR KIAA1506.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
TISSUE=Fetal thymus;
MEDLINE=21888622; PubMed=11891048;
RX Ruault M., Brun M.-E., Ventura M., Roizes G., De Sario A.;
RT "MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a
RT chromosome region frequently deleted in myeloid leukaemia.";
RL Gene 284:73-81(2002).
[2]
SEQUENCE FROM N.A. (ISOFORM 2).
TISSUE=Cervical carcinoma;
MEDLINE=21574953; PubMed=11718452;
RX Tan Y.C., Chow V.T.;
RA "Novel human HALR (MLL3) gene encodes a protein homologous to ALR and
RT to ALR-1 involved in leukemia, and maps to chromosome 7q36 associated
RT with leukemia and developmental defects.";
RL Cancer Detect. Prev. 25:454-469(2001).
[3]
SEQUENCE FROM N.A.
MEDLINE=22737999; PubMed=12853948;
RA Hiller L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,
RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,
RA Wyllie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,
RA Fowell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,
RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,
RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,
RA Orzelsky P., Bielicki L., Scott K., Holmes A., Harkins R., Harris A.,
RA Strong C.W., Hou S., Tomlinson C., Dauphin-Kohlberg S.,
RA Kozlowski-Reilly A., Leonard S., Rohlfing T., Rock S.M.,
RA Tin-Willam A.-M., Abbott A., Minx P., Maupin R., Strommatt C.,
RA Latreille P., Miller N., Johnson D., Murray J., Woessner J.P.,
RA Wendt M.C., Yang S.-P., Schultz B.R., Wallis J.W., Spieth J.,
RA Bieri T.A., Nelson J.O., Berkowicz N., Wohlmann P.E., Cook L.L.,
RA Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,
RA Clifton S.W., Chisoe S.B., Marra M.A., Raymond C., Haugen E.,
RA Gillett W., Zhou Y., James R., Phelps K., Iadonato S., Bubb K.,
RA Simms E., Levy R., Clendenning J., Kaul R., Kent W.J., Furey T.S.,
RA Baertsch R.A., Brent M.R., Keibler E., Flicek P., Bork P., Suyama M.,
RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,
RA Eddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,
RA Waterston R.H., Wilson R.K.;
RT "The DNA sequence of human chromosome 7.";
```

```

RL Nature 424:157-164(2003).
RN [4]
RP SEQUENCE OF 556-3865 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT the complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [5]
RP SEQUENCE OF 3193-3865 AND 4460-4911 FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 3879-4911 FROM N.A.
RC TISSUE=Testis;
RA Duesterhoeft A., Lauber J., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX (ISOFORM 2).
RC TISSUE=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968;
RA Goo Y.-H., Sohn Y.C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.-J., Lee K.-J., Lee Y.C.,
RA Lee J.W.;
RT "Activating signal cointegrator 2 belongs to a novel steady-state
RT complex that contains a subset of trithorax group proteins.";
RL Mol. Cell. Biol. 23:140-149(2003).
CC -!- FUNCTION: Belongs to the ASC-2/NCOA6 complex (ASCOM), a
CC coactivator complex of nuclear receptors, involved in
CC transcriptional coactivation. MLL3 may be a catalytic subunit of
CC this complex, which weakly methylates Lys-4 of histone H3. This is
CC a specific tag for epigenetic transcriptional activation. May be
CC involved in leukemogenesis and developmental disorder.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
CC -!- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins
CC MLL2 and MLL3, and ASH2/ASCL2. Interacts with histone H3.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name1:
CC IsoId=Q8NEZ4-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8NEZ4-2; Sequence=VSP_008561, VSP_008562;
CC -!- TISSUE SPECIFICITY: Highly expressed in testis and ovary, followed
CC by brain and liver. Also expressed in placenta, peripheral
CC blood, fetal thymus, heart, lung and kidney. Within brain,
CC expression was highest in hippocampus, caudate nucleus, and
CC substantia nigra. Not detected in skeletal muscle and fetal liver.
CC DOMAIN: The SET domain interacts with histone H3 but not H2A, H2B
CC and H4, and may have a H3 lysine specific methylation activity.
CC -!- MISCELLANEOUS: Found in a critical region of chromosome 7, which
CC is commonly deleted in malignant myeloid disorders. Partial
CC duplication of the MLL3 gene are found in the juxtacentromeric
CC region of chromosomes 1, 2, 13 and 21. Juxtacentromeric
CC reshuffling of the MLL3 gene has generated the BAGE genes.
CC -!- SIMILARITY: Belongs to the TRX/MLL family.
CC -!- SIMILARITY: Contains 1 DHHC-type zinc finger.
CC -!- SIMILARITY: Contains 6 PHD-type zinc fingers.
CC -!- SIMILARITY: Contains 1 post-SET domain.
```



CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
CC -!- SIMILARITY: Contains 1 SET domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF024361; AK00583.1; -;  
CC EMBL; AF264750; AAF74766.2; -;  
CC EMBL; AC006017; AAD45822.1; -;  
CC EMBL; AC104592; -; NOT ANNOTATED\_CDS.  
CC EMBL; AC005631; -; NOT ANNOTATED\_CDS.  
CC EMBL; AB040939; BA96030.2; -;  
CC EMBL; AK022687; BAB14179.1; -;  
CC EMBL; AK075113; BAC11409.1; -;  
CC EMBL; AL833924; CAD38780.1; -;  
CC Genew; HGNC:13726; MLL3.  
CC MIM; 606833; -;  
CC InterPro; IPR000637; AT\_hook.  
CC InterPro; IPR003889; FYRICH\_C.  
CC InterPro; IPR003888; FYRICH\_N.  
CC InterPro; IPR000910; HMG\_12\_box.  
CC InterPro; IPR003616; PostSET.  
CC InterPro; IPR001214; SET.  
CC InterPro; IPR001594; Znf\_DHHC.  
CC InterPro; IPR001965; Znf\_PHD.  
CC InterPro; IPR001841; Znf\_Ring.  
CC Pfam; PF00505; HMG\_box; 1.  
CC Pfam; PF00628; PHD; 6.  
CC Pfam; PF00856; SET; 1.  
CC SMART; SM00542; FYRC; 1.  
CC SMART; SM00541; FYRN; 1.  
CC SMART; SM00398; HMG; 1.  
CC SMART; SM00249; PHD; 8.  
CC SMART; SM00508; PostSET; 1.  
CC SMART; SM00317; SET; 1.  
CC PROSITE; PS00354; HMG1\_Y; 1.  
CC PROSITE; PS50868; POST\_SET; 1.  
CC PROSITE; PS50280; SET; 1.  
CC PROSITE; PS50216; ZF\_DHHC; 1.  
CC PROSITE; PS01359; ZF\_PHD; 5.  
CC PROSITE; PS50016; ZF\_PHD\_2; 6.  
CC PROSITE; PS50089; ZF\_RING\_2; 1.  
CC Transferrase; Methyltransferase; Chromatin regulator; Activator;  
CC DNA-binding; Nuclear protein; Transcription regulation; Coiled coil;  
CC Zinc-finger; Repeat; Alternative splicing; Polymorphism.  
CC ZN\_FING 341 391 PHD-TYPE 1.  
CC ZN\_FING 344 389 RING-TYPE.  
CC ZN\_FING 388 438 PHD-TYPE 2.  
CC ZN\_FING 436 489 DHHC-TYPE.  
CC ZN\_FING 464 520 PHD-TYPE 3.  
CC ZN\_FING 957 1010 PHD-TYPE 4.  
CC ZN\_FING 1007 1057 PHD-TYPE 5.  
CC ZN\_FING 1084 1139 PHD-TYPE 6.  
CC ZN\_FING 1470 4891 SET.  
CC DOMAIN 4770 4891 POST-SET.  
CC DOMAIN 4895 4911 COILED COIL (POTENTIAL).  
CC DOMAIN 92 112 COILED COIL (POTENTIAL).  
CC DOMAIN 644 672 COILED COIL (POTENTIAL).  
CC DOMAIN 1338 1366 COILED COIL (POTENTIAL).  
CC DOMAIN 1754 1787 COILED COIL (POTENTIAL).  
CC DOMAIN 3054 3081 COILED COIL (POTENTIAL).  
CC DOMAIN 3273 3272 COILED COIL (POTENTIAL).  
CC DOMAIN 3391 3433 COILED COIL (POTENTIAL).  
CC DOMAIN 34 46 A.T HOOK (BY SIMILARITY).  
CC DOMAIN 1719 1796 GLN-RICH.  
CC DOMAIN 1834 2281 PRO-RICH.  
CC DOMAIN 2412 2630 PRO-RICH.  
CC DOMAIN 2690 2786 ASP-RICH.

Query Match 9.5%; Score 136.5; DB 1; Length 4911;  
Best Local Similarity 24.0%; Pred. NO. 0.57;  
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;  
QY 29 ARTSYVDTEFLFGSPAGTR-----PTPPDFDPPWVK-----AMRTGVGK 68  
DB 1933 SRPLQWNETTANRSPVYDLCSSTTNDYFAYKPPDTPFVMTDQPFKSLGLSRSPVSE 1992  
QY 69 EASKALGAKGSC-----ETTPSRGSGTPTLTPRKVK-----YPISTHPS 108  
DB 1993 QTAGPIAAGTSDHFTKPSRADYVQRIQIPDSYARPLLPALDPSGFGFKTQMPPPS 2052  
QY 109 YCDESLSGSRSEGSFCAIPMAKGAALRALLTWPTPTPRGSH--SPRPREAPLRAIHP 166  
DB 2053 SQDP--YGSVSG-----ASRLSVDYPERPAL--TPRPIDNESHNQNDYSPQLTPHP 2103  
QY 167 -----AGPSK--TEPG-----PAASQKLXSMGLHS-----SRPL 194  
DB 2104 AVNESFAHPSRAFSQPGTISRPTSDQPSYQPPGPRPRVDSYSQSOSGTAASNTDPSQSP- 2162  
QY 195 KRGLSHSLTHLNVSTGHPATSAHTNGRDLRSTSGVTFRSPLVTSRARSVSISVEST 254  
DB 2163 -----FGTPRPTTVDSYQQQPPRSTQTDLFVTPVTNQRHSDPYAHPPGT 2209  
QY 255 PRGGATQKPKPP 267  
DB 2210 PRGISVPSQPP 2222  
RESULT 3  
OSA\_DROME  
ID OSA\_DROME STANDARD; PRT; 2716 AA.  
AC Q8IN94; O61603; Q9VEG7;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Trithorax group protein OSA (Eyelid protein).  
GN OSA OR ELD OR CG7467.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE  
RP SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RX MEDLINE=97415319; PubMed=9271118;  
TX Treisman J.E., Luk A., Rubin G.M., Heberlein U.;  
RT "eyelid antagonizes wingless signaling during Drosophila development  
RT and has homology to the Bright family of DNA-binding proteins."  
RL Genes Dev. 11:1949-1962(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brannon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fessler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,



QY 68 KEASKALGAKGSCETTPSRGSGTTLTPRKKNKYPISHTPSYCDSESLFSGRSEGASFGAP 127  
Db 689 GANSMSGAGAG--GYPGRG-----MPNHTGQYPPQWVPSPQQTVPQGGAGVGNH 741  
QY 128 RMAKGDAKRLALLWTP-----PPTPRGSHSPRREAPLRAIHAPGSKTEPQPAADS 180  
Db 742 VQKGG-----TPPPFVGVGGPPPGSGSPRELNYLKQHLQHKGGYGGSPTPPQGP 791  
QY 181 OKLSMG--GLHSSRP-----XKRLSHSLTHLNVPT-----GHPATSAPHT 220  
Db 792 QYGVNGPTGHPMPMGPHNGPPHGPINMGPTSTPPQSQVQLQGQPGQGGASGGRES 851  
QY 221 NGQDLRPSTGVTFRSPLVTSRARSVSISVSTP 255  
Db 852 GGSEHI-SQDNGISSGPTGAGMHAIVTSVVTGP 885

RESULT 4  
WAIP HUMAN STANDARD; PRT; 503 AA.  
ID WAIP HUMAN AC 043516; Q15220; Q9UNP1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Wiskott-Aldrich syndrome protein interacting protein (WASP) interacting  
DE protein) (PRL-2 protein).  
GN WASPIP OR WIP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CX NCBI\_TaxID=9606;  
[1]  
RN R  
RP MEDLINE=98070810; PubMed=9405671;  
RX Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;  
RA "WIP, a protein associated with Wiskott-Aldrich syndrome protein,  
RT induces actin polymerization and redistribution in lymphoid cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676 (1997).  
[2]  
RN R  
RP SEQUENCE FROM N.A.  
TX TISSUE=Tonsil;  
RA Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
[3]  
RN R  
RP SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.  
RX MEDLINE=9218549; PubMed=10202051;  
RA Stewart D.M., Tian L., Nelson D.L.;  
RT "Mutations that cause the Wiskott-Aldrich syndrome impair the  
RT interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP  
RT interacting protein.";  
RL J. Immunol. 162:5019-5024 (1999).  
CC -1- FUNCTION: May have direct activity on the actin cytoskeleton.  
CC -1- Induces actin polymerization and redistribution.  
CC -1- SUBUNIT: Binds to WASP, profilin and actin.  
CC -1- TISSUE SPECIFICITY: Highly expressed in peripheral blood  
CC mononuclear cells, spleen, placenta, small intestine, colon,  
CC thymus. Lower expression in ovary, heart, brain, lung, liver,  
CC skeletal muscle, kidney, pancreas, prostate and testis.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF031588; AAC03767.1; -;  
CC EMBL; AF106062; AAD45972.1; -;  
CC EMBL; X86019; CAA60014.1; -;  
CC Genew; HGNC:12736; WASPIP.  
CC MIM; 602357; -;

DR GO; GO:0015629; C:actin cytoskeleton; TAS.  
DR GO; GO:0003779; F:actin binding; TAS.  
DR GO; GO:0005522; F:profilin binding; TAS.  
DR GO; GO:0008154; P:actin polymerization and/or depolymerization; TAS.  
DR GO; GO:0006461; P:protein complex assembly; TAS.  
DR InterPro: IPR003124; WH2.  
DR Pfam; PF02205; WH2; 1.  
DR SMART; SMO0246; WH2; 1.  
KW Actin-binding; Repeat; Polymorphism.  
FT DOMAIN 2 13 POLY-PRO.  
FT DOMAIN 64 96 GLY-RICH.  
FT DOMAIN 241 244 POLY-SER.  
FT DOMAIN 362 433 PRO-RICH.  
FT REPEAT 352 361 XRSGLPXPX MOTIF 1.  
FT REPEAT 374 383 XRSGLPXPX MOTIF 2.  
FT REPEAT 410 419 XRSGLPXPX MOTIF 3.  
FT SITE 45 48 ACTIN BINDING.  
FT VARIANT 495 495 G -> A.  
FT /FTID=VAR\_010295.  
FT CONFLICT 303 309 PRRHLR -> SSOAPP (IN REF. 3).  
FT CONFLICT 360 360 P -> PV (IN REF. 2).  
FT CONFLICT 487 503 GSNRRERGGPPLPIPR -> EYFCQGF (IN REF. 2).  
SQ SEQUENCE 503 AA; 51489 MW; 43EB88674DD3BF1A CRC64;  
Query Match 9.4%; Score 134.5; DB 1; Length 503;  
Best Local Similarity 25.8%; Pred. No. 0.077;  
Matches 63; Conservative 23; Mismatches 107; Indels 51; Gaps 11;  
QY 42 PAGTR-PTPPDFDPWVEKANTRGVGKASKALGAKGSCETTPSRGSGTPT--LTPKKN 98  
Db 207 PGGPRQSPGPTPPPPFPGNRTALGGGSIROSPLSSSPFNRRPLPPTPSRALDDKPP 266  
QY 99 KYRPISTPSYCDSESLFSGRSEGASFGAPRMAKDAKRLALLWTPPTPRGSHSPRPR- 157  
Db 267 PPPVGNRPSTHRAV-----PPPPQNNKPPVPST 297  
QY 158 ---EAPLRA-IHPAGPSKTEPGPAADSKLNGGLHSSRLKRLSHSLTHLNVPTSGH- 212  
Db 298 PRPSAPHRPLRPPLPPSRPGPPPLPPS---SSGNDETPLPQLNLSLSSTPTLPSPGRS 354  
QY 213 -PATSAPHTNGFDLR--PSTSGVTFRSPLVTSRARSVSISVSTPSTP---PGGATQK--- 263  
Db 355 GLPPLPPPSERPPPPVDRDPFGRSG-PLPPPPVSRNGSTSRALPATPQLPSRSGVDSPRSG 413  
QY 264 PKPP 267  
Db 414 PRPP 417  
RESULT 5  
ID SYN1 HUMAN STANDARD; PRT; 705 AA.  
AC P17600; O75825;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Synapsin I (Brain protein 4.1).  
GN SYN1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CX NCBI\_TaxID=9606;  
[1]  
RN R  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=90243651; PubMed=2110562;  
RA Suedhof T.C.;  
RT "The structure of the human synapsin I gene and protein.";  
RL J. Biol. Chem. 265:7849-7852 (1990).  
[2]  
RN R  
RP SEQUENCE OF 1-125 FROM N.A.  
RX MEDLINE=90368667; PubMed=2118519;  
RA Sauerwald A., Hoesche C., Oschwald R., Killmann M.W.;



RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
RA Dami P.D., Dockree C., Dodsorth S.J., Durbin R.M., Ellington A.G.,  
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,  
RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,  
RA Hunt S.E., Jones M.C., Karshay J., Kimberley A.M., King A.,  
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
RA Martyn I.D., Mashreghi-Mohammadi A., Matthews L.H., Mccann O.T.,  
RA Mccloy J., McLaren S., McMuray A.A., Milne S.A., Mortimore B.J.,  
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,  
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuoyama S.,  
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Keston S., Lai H., Lao H.I.,  
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull B., Sloan D., Song L.,  
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,  
RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,  
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,  
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,  
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
RA Edelman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,  
RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
RA Khan A.S., Lane L., Tilahun Y., Wright H.,  
RA "The DNA sequence of human chromosome 22,"  
RA Nature 402:489-495(1999).  
[3]  
SEQUENCE OF 25-863 FROM N.A.  
TISSUE=Melanoma;  
RA Ansoorge W., Winkler U., Mewes H.-W., Weil B., Wiemann S.;  
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE OF 73-863 FROM N.A.  
TISSUE=Brain;  
RA Tissue=Brain;  
RA Hirotsawa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;  
RA "Identification of novel transcribed sequences on human chromosome 22  
RA by expressed sequence tag mapping."  
RA DNA Res. 8:1-9(2001).  
[5]  
SEQUENCE OF 514-863 FROM N.A.  
TISSUE=Choriocarcinoma;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalhus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RA "Generation and initial analysis of more than 15,000 full-length human  
RA and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC - SUBUNIT: Binds to Rab13.  
CC - SIMILARITY: Contains 1 calponin-homology (CH) domain.  
CC - SIMILARITY: Contains 1 LIM zinc-binding domain. that shown due to  
CC - CAUTION: Ref.3 (CAD39036) sequence differs from that shown due to  
CC frameshifts in positions 486 and 507.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ496196; CAD42713.1; -  
CC EMBL; AL022311; CAB62980.1; -  
CC EMBL; AL022311; CAB62979.1; -  
CC EMBL; AL022311; CAB62978.1; ALT\_SEQ.  
CC EMBL; AL834373; CAD39036.1; ALT\_FRAME.  
CC EMBL; AL833860; CAD38718.1; -  
CC EMBL; AB051455; BAB33338.1; -  
CC EMBL; BC001090; AAH01090.2; -  
CC HSSP; Q01082; 1BKE.  
CC GO; GO:0005622; C:intracellular; IC.  
CC GO; GO:0016192; P:vesicle-mediated transport; IC.  
CC InterPro; IPR001715; Calponin-like.  
CC InterPro; IPR001781; LIM.  
CC Pfam; PF00307; CH; 1.  
CC Pfam; PF00412; LIM; 1.  
CC ProDom; PD000094; LIM; 1.  
CC SMART; SM00033; CH; 1.  
CC SMART; SM00132; LIM; 1.  
CC PROSITE; PS00021; CH; 1.  
CC PROSITE; PS00478; LIM DOMAIN 1; 1.  
CC PROSITE; PS00023; LIM DOMAIN 2; 1.  
KW LIM domain; Metal-binding; Zinc.  
FT DOMAIN 2 105  
FT DOMAIN 162 225  
FT DOMAIN 253 667  
FT DOMAIN 428 437  
FT DOMAIN 137 137  
FT CONFLICT 212 212 P -> Y (IN REF. 3; CAD39036).  
FT CONFLICT 519 519 A -> S (IN REF. 3).  
SQ SEQUENCE 863 AA; 93441 MW; 755E3B57C6037292 CRC64;  
  
Query Match 9.0%; Score 129.5; DB 1; Length 863;  
Best Local Similarity 25.9%; Pred. No. 0.27;  
Matches 64; Conservative 30; Mismatches 130; Indels 23; Gaps 11;  
  
QY 40 GSPAGTRTP-PDPDPW---VEKANRTGVGKESKALGAKGSCETTPSRGST----P 90  
DB 354 GTPKPSGTPAPRKPPWITLVQAEPKKPPAPLPFSSGFPSPQSRQVENGTEVAQP 413  
QY 91 TLTPKKNK-YRPIGHTSPSYCDSEIFGSRSEGAFFGAPFMAKGAALRLW---TPPP 146  
DB 414 SPTASLESKPYNPFEEEDKEEAPAPSLATS---PALGHPESTPKSLHPWYGITPTS 470  
QY 147 TTRGSHSPRPR---EAPLAIHPAGSKTEPGPADSQKLSMGLHSSRPKKGLSHSLT 203  
DB 471 SPKTKRPAAPRASPL-ALFASRLSHSEPPSATPSPALSVESLSSESASQTGAELLE 529  
QY 204 HLNVP-STCHPATSAPHTNGPQDLRPSVTSGVTFRSPLTSTRARSVISVPS-TPR-RGGA 260  
DB 530 PPAVPSKSEPAVHAPGTGPNVPLSTNSLASSSELVEPRVEQMPQASPLAPTRGSS 589  
QY 261 TQPKPFP 267  
DB 590 GPQAPKE 596  
  
RESULT 7  
P121\_RAT

ID P121 RAT STANDARD; PRT; 1199 AA.  
AC P52591;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein of 121 kDa) (P145).  
DE of 121 kDa) (P145).  
GN POM121 OR NUP121.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=Sprague-Dawley; Tissue=Liver;  
RX MEDLINE=93328754; PubMed=9335683;  
RA Hallberg B.; Wozniak R.W.; Blobel G.;  
RT "An integral membrane protein of the pore membrane domain of the nuclear envelope contains a nucleoporin-like region."  
RL J. Cell Biol. 122:513-521(1993).  
CC -!- FUNCTION: Essential component of the nuclear pore complex. The repeat-containing domain may be involved in anchoring components of the pore complex to the pore membrane.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE MEMBRANE.  
CC -!- DOMAIN: Contains P-X-F-G repeats.  
CC -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z21513; CAA79725.1; -;  
CC EMBL; Z21514; CAA79726.1; -;  
CC PIR; A40670; A40670.  
CC InterPro; IPR004325; Nucleoporin FG.  
CC Pfam; PF03093; Nucleoporin FG; 8.  
CC Nucleoporin; Transport; Transmembrane; Repeat.  
CC DOMAIN 1 56 CISTERNAL SIDE (POTENTIAL).  
CC TRANSMEM 57 77 POTENTIAL.  
CC DOMAIN 76 1199 PORE SIDE (POTENTIAL).  
CC DOMAIN 4 8 POLY-ALA.  
CC DOMAIN 53 58 POLY-ALA.  
CC DOMAIN 438 441 POLY-SER.  
CC DOMAIN 497 500 POLY-PRO.  
CC DOMAIN 684 687 POLY-SER.  
CC DOMAIN 1023 1026 POLY-ALA.  
CC DOMAIN 1033 1038 POLY-SER.  
CC DOMAIN 1058 1063 POLY-GLY.  
CC SEQUENCE 1199 AA; 120784 MW; 6DC4451B91D5B907 CRC64;  
CC  
CC Query Match 8.8%; Score 126.5; DB 1; Length 1199;  
CC Best Local Similarity 26.1%; Pred. No. 0.59;  
CC Matches 67; Conservative 26; Mismatches 131; Indels 33; Gaps 11;  
CC  
CC 28 KARTSVYDELTF---GSPAGTRTPDPDPVPEXANTRGVGKSKALGKSCETT 83  
CC 362 RSRITSSVSLTSTCTGIPSSRNA-----ITSSVSTRGV-SQWKGSGTSPFSS 413  
CC 84 P--SRGTPPLTRPKNKVRPISTHTSYCDSELSFGSRSEGAFGRMAKGAADAKRL 141  
CC 414 PASRSQTPE-RAKKTREEPCQSSSAPLVTDKESPCEKVTDPATGKQSS-----L 466  
CC  
CC 142 WTPPPPTGRGSHSPRPEAPL---RAHPAGRSKTEFGPAADSOKLSMGLHSSRLKRG 197  
CC 467 WTSPTPT-GSSGQKRXKIQLLPSRRGDQLTLPPLPELGISITAEDLDMERRASLQWFKV 525

QY 198 LSHSLTHLNVSTG-HPATSAHTNGPODLRSTSGVTF-----RSPLVTSRARSVSI 252  
DB 526 LEDKTDASTPATDTSPTSPFTLTPTVGFAASPASPSPNPLLESLSKMQQESAP 585  
QY 253 ST--PRRGATQKPKPP 267  
DB 586 SSSEPEAATVAARSP 602  
  
RESULT 8  
IRS2 MOUSE  
ID IRS2 MOUSE STANDARD; PRT; 1321 AA.  
AC P81122;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Insulin receptor substrate-2 (IRS-2) (4FS).  
DE Insulin receptor substrate-2 (IRS-2) (4FS).  
GN IRS2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=95405472; PubMed=7675087;  
RA Sun X.J.; Wang L.-M.; Zhang Y.; Yenush L.; Myers M.G. Jr.;  
RA Glasheen E.; Lane W.S.; Pierce J.H.; White M.F.;  
RT "Role of IRS-2 in insulin and cytokine signalling."  
RL Nature 377:173-177(1995).  
CC -!- FUNCTION: May mediate the control of various cellular processes by insulin.  
CC -!- TISSUE SPECIFICITY: Skeletal muscle, lung, brain, liver, kidney, heart and spleen.  
CC -!- SIMILARITY: Contains 1 PH domain.  
CC -!- SIMILARITY: Contains 1 PTB domain.  
CC HSSP; P35568; 1IRS.  
CC MGD; MGI:109334; IRS2.  
CC InterPro; IPR002404; Insln\_receptorS1.  
CC InterPro; IPR001849; PH.  
CC Pfam; PF02174; IRS; 1.  
CC Pfam; PF00169; PH; 1.  
CC PRINTS; PR00528; INSULINRS1.  
CC SMART; SM00233; PH; 1.  
CC SMART; SM00310; PTB; 1.  
CC PROSITE; PS00003; PH\_DOMAIN; 1.  
CC Phosphorylation.  
CC KW DOMAIN 16 144 PH.  
CC FT DOMAIN 187 300 PTB.  
CC FT MOD\_RES 536 536 PHOSPHORYLATION (BY INSR)  
CC FT MOD\_RES 649 649 PHOSPHORYLATION (BY INSR)  
CC FT MOD\_RES 671 671 PHOSPHORYLATION (BY INSR)  
CC FT MOD\_RES 911 911 PHOSPHORYLATION (BY INSR)  
CC FT MOD\_RES 970 970 PHOSPHORYLATION (BY INSR)  
CC FT MOD\_RES 1242 1242 PHOSPHORYLATION (BY INSR)  
CC FT MOD\_RES 1303 1303 PHOSPHORYLATION (BY INSR)  
CC FT DOMAIN 19 28 POLY-ASN.  
CC FT DOMAIN 444 449 POLY-SER.  
CC FT DOMAIN 638 641 POLY-SER.  
CC FT DOMAIN 936 939 POLY-SER.  
CC SQ SEQUENCE 1321 AA; 136526 MW; 5069CB9D614960C7 CRC64;  
  
Query Match 8.8%; Score 126.5; DB 1; Length 1321;  
Best Local Similarity 26.5%; Pred. No. 0.65;  
Matches 73; Conservative 31; Mismatches 132; Indels 39; Gaps 12;  
  
QY 4 PVELAVSGMOTLGLQHRRCGGYRVKARTSYVDTLFGSPAGTRTPDPDPVPEXANRT 63

Db 882 PTRLSLEGLTLPMSQEPYLPTEPKSPGEVIN--IDPGEAGTSLSP--APLLASAASS 937  
 Qy 64 RGVGKAKA--LGAAGSCETTSPRGSTPTLPTRKKNKVPISHTSYCDESLFGSR--- 118  
 Db 938 SLLSASSPASSLGSCTPGTSSDRQRSPSLDYMNLDPSPKPKSTSGTSGTVGSMOGL 997  
 Qy 119 -SEGAFGAPRMAKGAALRALLWTPTPPRGSHSPRPREAPLRAIHPAGPS--KTEP 174  
 Db 998 LSPERS--SP-----YPLPFRSTSPSLQOPL---PPAFGLYRLPP 1036  
 Qy 175 GPAADSOGLSMGLHSGRLPKRGLSHSLTHLNVST-GHPATSAHTNGPQDLRPSTGV 233  
 Db 1037 ASAATSGPTAGSMSEPDNG-DYSEMAFGVAATPPQIVAPPKPEGARVAP--TSGL 1094  
 Qy 234 TFSRPL--VTSRARSVISVSPTRRGATQKPKP 266  
 Db 1095 KRUSLMDQVSGVEAFICVSPDPHKGAKVIRADP 1129

## RESULT 9

SYN1\_BOVIN  
 ID SYN1\_BOVIN STANDARD; PRT; 706 AA.  
 AC P17599;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Synapsin I.  
 GN SYN1  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain.  
 RX MEDLINE=89388265; PubMed=2506642;  
 RA Suedhof T.C., Czernik A.J., Kao H.-T., Takei K., Johnston P.A.,  
 RA Horiuchi A., Kanazir S.D., Wagner M.A., Perin M.S., de Camilli P.,  
 RA Greengard P.;  
 RT "Synapsins: mosaics of shared and individual domains in a family of  
 RT synaptic vesicle phosphoproteins";  
 RL Science 245:1474-1480(1989).  
 RN [2]  
 RP PHOSPHORYLATION SITE BY PROLINE-DIRECTED KINASE.  
 RX MEDLINE=90216728; PubMed=2108963;  
 RA Hall F.L., Mitchell J.P., Vulliamt P.R.;  
 RA "Phosphorylation of synapsin I at a novel site by proline-directed  
 RT protein kinase";  
 RL J. Biol. Chem. 265:6944-6948(1990).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 112-417.  
 RX MEDLINE=98130589; PubMed=9463376;  
 RA Esser L., Wang C.R., Hosaka M., Smagula C.S., Suedhof T.C.,  
 RA Deisenhofer J.;  
 RT "Synapsin I is structurally similar to ATP-utilizing enzymes";  
 RL EMBO J. 17:977-984(1998).  
 CC -!- FUNCTION: Neuronal phosphoprotein that coats synaptic vesicles,  
 CC binds to the cytoskeleton, and is believed to function in the  
 CC regulation of neurotransmitter release.  
 CC -!- SUBCELLULAR LOCATION: Synapse.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=Synapsin IA;  
 CC IsoId=P17599-1; Sequence=Displayed;  
 CC Name=Synapsin IB;  
 CC IsoId=P17599-2; Sequence=VSP\_006314; VSP\_006315;  
 CC -!- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS  
 CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF  
 CC SYNAPSIN I IN THE NERVE TERMINAL.  
 CC -!- SIMILARITY: Belongs to the synapsin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M27810; AAA30761.1; -;  
 DR EMBL; M27811; AAA30762.1; -;  
 DR PDB; 1AUV; 18-MAR-98.  
 DR PDB; 1AUX; 18-MAR-98.  
 DR GlycoSuiteDB; P17599;  
 DR InterPro; IPR001359; Synapsin.  
 DR Pfam; PF02078; Synapsin\_1.  
 DR Pfam; PF02750; Synapsin\_C; 1.  
 DR PRINTS; PR01368; SYNAPSIN.  
 DR PROSITE; PS00415; SYNAPSIN\_1; 1.  
 DR PROSITE; PS00416; SYNAPSIN\_2; 1.  
 KW Synapse; Phosphorylation; Neurone; Repeat; Actin-binding;  
 KW Alternative splicing; 3D-structure.  
 FT DOMAIN 1 28 A.  
 FT DOMAIN 29 112 B (LINKER).  
 FT DOMAIN 113 420 C  
 FT (ACTIN-BINDING AND SYNAPTIC-VESICLE  
 FT BINDING).  
 FT D (PRO-RICH LINKER).  
 FT E.  
 FT PHOSPHORYLATION (BY PKA AND CAMK1).  
 FT PHOSPHORYLATION (BY PDPK).  
 FT PHOSPHORYLATION (BY CAMK2).  
 FT PHOSPHORYLATION (BY CAMK2).  
 FT NKSQLTNA -> KASPAQAQP (in isoform  
 FT Synapsin IB).  
 FT /FTID=VSP\_006314.  
 FT Missing (in isoform Synapsin IB).  
 FT /FTID=VSP\_006315.  
 FT  
 FT VARSPLIC 671 706  
 FT STRAND 114 119  
 FT TURN 122 123  
 FT HELIX 126 130  
 FT TURN 131 132  
 FT STRAND 134 135  
 FT TURN 136 138  
 FT STRAND 139 146  
 FT TURN 149 150  
 FT STRAND 151 155  
 FT TURN 157 158  
 FT STRAND 161 165  
 FT TURN 172 173  
 FT STRAND 176 178  
 FT STRAND 182 185  
 FT STRAND 190 190  
 FT TURN 194 195  
 FT STRAND 197 197  
 FT HELIX 199 207  
 FT TURN 208 209  
 FT STRAND 212 213  
 FT HELIX 216 221  
 FT TURN 222 223  
 FT HELIX 225 229  
 FT TURN 241 243  
 FT STRAND 246 246  
 FT STRAND 250 252  
 FT HELIX 255 257  
 FT STRAND 266 270  
 FT TURN 275 276  
 FT STRAND 279 282  
 FT HELIX 285 296  
 FT TURN 297 299  
 FT STRAND 302 306  
 FT STRAND 310 319  
 FT TURN 320 321  
 FT STRAND 322 327





CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; L07922; AAA02783.1; -;  
DR EMBL; L07923; AAA02787.1; -;  
DR EMBL; X67776; CAA47986.1; -;  
DR PIR; S29605; S29605.  
DR InterPro; IPR007796; Herpes\_BULF1; 1.  
DR Membrane; Glycoprotein; Antigen; Late protein.  
KW CARBOHYD 47 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 169 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 195 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 229 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 328 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 378 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 386 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 411 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 435 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 443 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 457 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 497 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 519 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 533 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 568 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 589 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 624 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 680 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 886 AA; 92388 MW; 4394F6130DCC8A CRC64;

Query Match 8.6%; Score 123; DB 1; Length 886;  
Best Local Similarity 23.9%; Pred. No. 0.72;  
Matches 79; Conservative 30; Mismatches 120; Indels 102; Gaps 15;

Qy 3 TPELVAYSGMOTGLQHCRCRGYVVKARTSVYDTELFGSPAGTRPTP-----PDFDPPW 56  
Db 529 TPTNATP-----PTLG-----KTSTSVATTPFNATSPITPAVTPPTNATPT 573  
Qy 57 VEKANRTG-----VGKEASKA-----LGAKSCB-----TTPSRGSTPTLTPRK 96  
Db 574 LGKTSPTSAVTTPTNATSPVTGTSPOANTNHTLGGTSTPTVTPSPKATSAVTTGQ 633  
Qy 97 KN-----KYRPISTHPSYCDLFGSRSEGFAGPRMAKG-----DAAKLRAL 140  
Db 634 HNTTSSTSNLSRPSISSETLSTSDNSTSHMPLTSAHTPGENITQVTPASTSTHH 693  
Qy 141 LWTPTPTPR-----GSHSRPRPREA-----PLEAHPAGFS-----KTE 173  
Db 694 VSTSSPAPRPTTQASGPGNSSTSTKPGVNVTKGPPKPNATSPQAPSGQKTAVPVTS 753  
Qy 174 PGPAADSOGLSMGLHSRPLKGLSHLTHLVNVPST---GHPATSAFPHNGQDLRPS 230  
Db 754 TGGKANS---TTGGKHT-----GHGARTSTSTFTDYGGDSTTPTRYNATLYPPT 803  
Qy 231 SGV-----TPRSPLVTSRARSVSVSPSTPR 256  
Db 804 SSKLRPRWTFSPPTTAQATVPVPPISQPR 834

RESULT 12  
VGP3\_EBV

ID VGP3\_EBV STANDARD; PRT; 907 AA.  
AC P03200; P03201;  
DT 21-JUN-1986 (Rel. 01, Created)  
DT 21-JUN-1986 (Rel. 01, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Envelope glycoprotein GP340/GP220 (Membrane antigen) (MA).  
GN BULF1.  
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Lymphocryptovirus.  
OX NCBI\_TaxID=10377;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS GP220 AND GP340).  
RX MEDLINE=84270667; PubMed=6087149;  
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,  
RA Tufnell P.S., Barrett B.G.;  
RL "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";  
RL Nature 310:207-211(1984).  
CC -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN  
CC B-CELLS.  
CC -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL  
CC ENVELOPE.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=GP340;  
CC IsoId=P03200-1; Sequence=Displayed;  
CC Name=GP220;  
CC IsoId=P03200-2; Sequence=VSP\_002070;  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; V01555; CAA24854.1; -;  
DR PIR; A43042; QQBE21.  
DR InterPro; IPR007796; Herpes\_BULF1; 1.  
DR Pfam; PF05109; Herpes\_BULF1; 1.  
KW Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.  
FT CARBOHYD 47 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 169 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 195 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 229 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 328 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 378 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 386 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 411 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 435 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 443 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 457 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 497 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 519 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 533 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 568 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 589 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 610 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 624 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 627 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 656 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 683 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 701 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 735 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 746 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 755 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 780 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 815 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 858 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 888 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 502 Missing (in isoform GP220).  
 FT /FTid=VSP\_002070.  
 SQ SEQUENCE 907 AA; 94431 MW; 0750141CEAC529 CRC64;  
 Query Match 8.5%; Score 122.5; DB 1; Length 907;  
 Best Local Similarity 22.8%; Pred. No. 0.79;  
 Matches 77; Conservative 24; Mismatches 128; Indels 109; Gaps 13;  
 QY 3 TPVELAVSGMOTGLQHCRCRGYRVKARTSVVDTLFGSPAGTRPT-----P 49  
 Db 543 TPTNATY--PTLG-----KTSPTSAVTTPTPNATSPILGKTSPTS AVTTPT 587  
 QY 50 PDFDPVWEKANRTGVGKEASKALG-----AKGSCETTPSRGSTTLTPRKKNKY 100  
 Db 588 PNATSPILGKTSPTS AVTTPTPNATGTVGTSPOANATNHTLGGTSPTPVTSQPNAT 647  
 QY 101 RPI-----SHTPSVCDSELSFGSRSEGASFGAPRMKAGDAAKRLALLWTPP 145  
 Db 648 SAVTTGQHNITSSSTSMRLPSSNPETLSPFSDNSTSMPLLTSAHTPGENITQVTP 707  
 QY 146 -----PTPR-----GSHSPREPREA-----PLRAHPAGPS----- 170  
 Db 708 ASISTHVSSTSPAPRGTTTSQASGPNSTSTKPGEVNVTGTPQONATSPQAPSGQKT 767  
 QY 171 -----KTEPGPADSQKLSMGLHSSRLPKLGLSHLTHLVNPT---GHPATSAHTNGP 223  
 Db 768 AVPTVTSTGGKANS---TTGGKHT-----GHGARTSTETPTDYGSDSTTPRPNAT 817  
 QY 224 QDLRPSGVS-----TFERSPLVTSRARSVSISVPSTPR 256  
 Db 818 TYLPPTSSKLRPWTFSPVTTAQTAVPVPPTSQPR 855  
 RESULT 13  
 CCAA\_MOUSE  
 ID\_CCAA\_MOUSE STANDARD; PRT; 2164 AA.  
 AC P97435;  
 DT 15-JUL-1999 (Rel. 38, Last Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium  
 DE channel, I type, alpha-1 polypeptide isoform 4) (Brain calcium  
 DE channel I) (BI).  
 GN CACNALA OR CACNL1A4 OR CCHAL1A OR CACH4 OR CACN3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANT TG LEU-601.  
 RC STRAIN=DNA/2J;  
 RX MEDLINE=97083572; PubMed=8929530;  
 RA Fletcher C.F., Lutz C.M., O'Sullivan T.N., Shaughnessy J.D. Jr.,  
 RA Hawkes R., Frankel W.N., Copeland N.G., Jenkins N.A.;  
 RT "Absence of epilepsy in tottering mutant mice is associated with calcium  
 RT channel defects."  
 RL Cell 87:607-617(1996).  
 CC -!- FUNCTION: Voltage-sensitive calcium channels (VSCC) mediate the  
 CC entry of calcium ions into excitable cells and are also involved  
 CC in a variety of calcium-dependent processes, including muscle  
 CC contraction, hormone or neurotransmitter release, gene expression,  
 CC cell motility, cell division and cell death. The isoform alpha-1A  
 CC gives rise to P and/or Q-type calcium currents. P/Q-type calcium  
 CC channels belong to the "high-voltage activated" (HVA) group and

are blocked by the funnel toxin (Ftx) and by the omega-agatoxin-  
 IVA (omega-Aga-IVA). They are however insensitive to  
 dihydropyridines (DHP), and omega-conotoxin-GVIA (omega-CTX-GVIA).  
 -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT  
 COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS  
 IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-  
 FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS  
 SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM  
 CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA  
 LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.  
 -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 -!- TISSUE SPECIFICITY: Brain specific; mainly found in the  
 cerebellum, olfactory bulb, cerebral cortex, hippocampus, and  
 inferior colliculus. In the hippocampus, expression occurs in  
 pyramidal and granule neurons, as well as in interneurons.  
 Purkinje cells contain predominantly P-type VSCC, the Q-type being  
 a prominent calcium current in cerebellar granule cells.  
 -!- DOMAIN: Each of the four internal repeats contains five  
 hydrophobic transmembrane segments (S1, S2, S3, S5, S6) and one  
 positively charged transmembrane segment (S4). S4 segments  
 probably represent the voltage-sensor and are characterized by a  
 series of positively charged amino acids at every third position.  
 -!- DISEASE: DEFECTS IN CACNALA ARE THE CAUSE OF A DELAYED-ONSET,  
 REGRESSIVE NEUROLOGICAL DISORDER SEEN IN TOTTERING (TG) MUTANTS,  
 RESULTING IN ATAXIA, MOTOR SEIZURES AND BEHAVIORAL ABSENCE  
 SEIZURES RESEMBLING PETIT MAL EPILEPSY (OR ABSENCE EPILEPSY) IN  
 HUMANS. THERE ARE TWO MORE ALLELES: LEANER (TG(LA)), THAT IS  
 CHARACTERIZED BY SEVERE ATAXIA AND FREQUENT DEATH PAST WEANING,  
 BUT NO MOTOR SEIZURES; AND ROLLING NAGAYA (TG(ROL)), THAT PRESENTS  
 AN INTERMEDIARY PHENOTYPE, THE ATAXIA BEING SOMEWHAT MORE SEVERE  
 THAT WITH TG, BUT WITHOUT MOTORS SEIZURES. SELECTIVE DEGENERATION  
 OF CEREBEULAR PURKINJE CELLS HAS BEEN SHOWN FOR ALL THESE TYPES OF  
 MUTANTS.  
 -!- SIMILARITY: Belongs to the calcium channel alpha-1 subunits  
 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U76716; AAC52940.1; -  
 MGD; MGI:109482; Cnacnla.  
 GO; GO:0005891; C:voltage-gated calcium channel complex; IC.  
 InterPro; IPR001682; Ca/Na pore.  
 InterPro; IPR002077; Ca channel alpha.  
 InterPro; IPR002111; Cat channel\_TrypL.  
 InterPro; IPR005821; Ion\_trans.  
 InterPro; IPR005820; M-channel\_nlg.  
 Pfam; PF00520; Ion\_trans; 4.  
 PRINTS; PR0167; CCHANNEL.  
 PRINTS; PR01632; PVDCCALPHA.  
 Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
 Calcium channel; Glycoprotein; Repeat; Multigene family;  
 Calcium-binding; Phosphorylation; Disease mutation.  
 REPEAT 17 317 I.  
 REPEAT 427 671 II.  
 REPEAT 1134 1417 III.  
 REPEAT 1454 1717 IV.  
 DOMAIN 1 52  
 TRANSMEM 53 71  
 DOMAIN 72 90  
 TRANSMEM 91 108  
 DOMAIN 109 120  
 TRANSMEM 121 136  
 DOMAIN 137 144  
 TRANSMEM 145 163  
 DOMAIN 164 182  
 TRANSMEM 183 202  
 CYTOPLASMIC (POTENTIAL).  
 S1 OF REPEAT I (POTENTIAL).  
 EXTRACELLULAR (POTENTIAL).  
 S2 OF REPEAT I (POTENTIAL).  
 CYTOPLASMIC (POTENTIAL).  
 S3 OF REPEAT I (POTENTIAL).  
 EXTRACELLULAR (POTENTIAL).  
 S4 OF REPEAT I (POTENTIAL).  
 CYTOPLASMIC (POTENTIAL).  
 S5 OF REPEAT I (POTENTIAL).

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 12:07:27 ; Search time 39 Seconds  
(without alignments)  
2176.267 Million cell updates/sec

Title: US-10-031-589-4  
Perfect score: 1437  
Sequence: 1 MKTPVELAVSGMOTLGLQHR.....SVFSTPRGGATQKPPPWK 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1437	100.0	269	4	Q96K30	Q96K30 homo sapien
2	867.5	60.4	253	11	Q9DJH0	Q9DJH0 mus musculus
3	863	60.1	177	4	Q96K25	Q96K25 homo sapien
4	150	10.4	647	10	Q8S148	Q8S148 oryza sativ
5	144.5	10.1	518	5	Q8MQG8	Q8MQG8 caenorhabdi
6	144.5	10.1	524	5	O02123	O02123 caenorhabdi
7	144.5	10.1	539	5	Q8MQG9	Q8MQG9 caenorhabdi
8	142.5	9.9	964	5	Q26963	Q26963 trypanosoma
9	136	9.5	769	10	Q9LLJ1	Q9LLJ1 chlamydomon
10	134	9.3	1131	5	Q9VLW2	Q9VLW2 drosophila
11	133.5	9.3	1303	13	Q7SV42	Q7SV42 brachydanio
12	133	9.3	862	11	Q9JTK1	Q9JTK1 rattus norv
13	133	9.3	1240	10	Q9DWH8	Q9DWH8 rat cytomeg
14	132.5	9.2	488	16	C87W07	C87W07 pseudomonas
15	132.5	9.2	506	12	C69125	C69125 human herpe
16	132.5	9.2	580	4	Q9UF83	Q9UF83 homo sapien

Q7t5d9 simian herp  
Q9Q0b8 herpes simp  
Q9Q0b7 herpes simp  
Q7f591 simian herp  
Q873d8 neurospora  
Q8azk7 human herpe  
Q8uzel cercopthic  
Q9fu38 oryza sativ  
Q9Q0b5 herpes simp  
Q7xss2 oryza sativ  
Q9Q0b4 herpes simp  
Q9Q0b3 herpes simp  
Q9Q0b9 herpes simp  
Q9Q0b2 herpes simp  
Q9Q0b6 herpes simp  
Q62185 caenorhabdi  
Q9ayf5 oryza sativ  
Q7xf26 oryza sativ  
Q9gzl1 caenorhabdi  
Q9Q0c1 herpes simp  
Q949f2 oryza sativ  
Q9be18 macaca fasc  
Q8rpv2 streptococc  
Q96dg4 homo sapien  
Q8av99 homo sapien  
Q8Q0c7 herpes simp  
Q9Q0c8 herpes simp  
Q9qcx4 chayote mos  
Q9hfi9 neurospora

ALIGNMENTS

RESULT 1

Q96K30	Q96K30	PRELIMINARY;	PRT;	269 AA.
AC	Q96K30;			
DT	01-DEC-2001 (TremBLrel. 19, Created)			
DT	01-DEC-2001 (TremBLrel. 19, Last sequence update)			
DT	01-OCT-2002 (TremBLrel. 22, Last annotation update)			
DE	Hypothetical protein FLJ14827 (Hypothetical protein			
DE	NT2RP1001002)			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,			
RA	Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,			
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,			
RA	Kinomiya K., Iwayanagi T.,			
RT	"NEDO human cDNA sequencing project."			
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RA	Strausberg R.,			
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,			
RA	Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,			
RA	Nagahari K., Sugano S., Isoigai T.,			
RT	"HRI human cDNA sequencing project."			
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AK027733; BAB55328.1; -			
DR	EMBL; BC022092; AAR2092.1; -			

```

DR EMBL; AK075358; BAC11568.1; -.
KW Hypothetical protein.
SQ SEQUENCE 269 AA; 28619 MW; 3F053BE1454F60773 CRC64;

Query Match      100.0%; Score 1437; DB 4; Length 269;
Best Local Similarity 100.0%; Pred. No. 5.5e-107;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFPPWVEKA 60
DB 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFPPWVEKA 60
QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120
DB 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120
QY 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSPRPREAPLRAIHPAGPSKTEPGGAADS 180
DB 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSPRPREAPLRAIHPAGPSKTEPGGAADS 180
QY 181 QKLSMGGHLSRPLKRGHSLHSLTHLNVSTGHPATSPHTNGPQDLRPSTSGVTFRSPLV 240
DB 181 QKLSMGGHLSRPLKRGHSLHSLTHLNVSTGHPATSPHTNGPQDLRPSTSGVTFRSPLV 240
QY 241 TSARSVSISVPSTPRGGATQKPPWK 269
DB 241 TSARSVSISVPSTPRGGATQKPPWK 269

RESULT 2
Q9DIH0 PRELIMINARY; PRT; 253 AA.
AC Q9DIH0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 110008J03Rik protein (RIKEN CDNA 110008J03 Gene).
GN 110008J03Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boiffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK003581; BAB22871.1; -.
DR EMBL; BC021365; AAB21365.1; -.

```

```

DR MGD; MG1:1922021; 1110008J03Rik.
SQ SEQUENCE 253 AA; 27070 MW; 239EC9BD9CF4FA61 CRC64;

Query Match      60.4%; Score 867.5; DB 11; Length 253;
Best Local Similarity 66.3%; Pred. No. 1.5e-61;
Matches 171; Conservative 15; Mismatches 67; Indels 5; Gaps 1;

QY 12 MOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFPPWVEKANTRGVGKEAS 71
DB 1 MKALHLQHRSPSYRVKARASVYVDETLFGSPARTRPAQDFDPFPMVQVNCNRSRGVGP 60
QY 72 KALGAKGSCETTPSRGSTPTLTPRKKNKYRPISTHTPSYCDLSFGSRSEGFAPGAPMAK 131
DB 61 KGSIAKRDCESSPSRGSTPNLTTPRKKNKYRILGHAPSVCDESLEFGTSKEGS-----RMAV 115
QY 132 GDAALRLALLWTPPTPRGSHSPRPREAPLRAIHPAGPSKTEPGGAADSQKLSMGGHLS 191
DB 116 GDAALRLFWTPPTPRGSHSPRPREAPLRAIHPAGPSKTEPGGAADSQKLSMGGHLS 175
QY 192 RPLKEGLSHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSL 251
DB 176 CSLGQRSSHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSLHSL 235
QY 252 PSTPRGGATQKPPWK 269
DB 236 PAPPRRGACPKPKPPWK 253

RESULT 3
Q96K25 PRELIMINARY; PRT; 177 AA.
AC Q96K25;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ14835.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovarian carcinoma;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sato H., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027741; BAB55333.1; -.
KW Hypothetical protein.
SQ SEQUENCE 177 AA; 19050 MW; 834FD0C7095AED0B CRC64;

Query Match      60.1%; Score 863; DB 4; Length 177;
Best Local Similarity 99.4%; Pred. No. 2.3e-61;
Matches 160; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFPPWVEKA 60
DB 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFPPWVEKA 60
QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120
DB 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120
QY 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSPRPREAPL 161
DB 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSPRPREAPV 161

```

```

RESULT 4
ID Q8S148 PRELIMINARY; PRT; 647 AA.
AC Q8S148;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE vSsa-like protein.
DE P0042A10.17.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
CX NCBI_TaxID=33947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RC Sasaki T., Matsumoto T., Yamamoto K.;
RR "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
RT clone:P0042A10."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003343; BAB90075.1; -.
DR Gramene; Q8S148; -.
DR InterPro; IPR002965; P rich extensin.
DR PRINTS; PR01217; PRICEXTNSN.
DR SQ SEQUENCE 647 AA; C27F3FACA77D1EC3 CRC64;

Query Match 10.48; Score 150; DB 10; Length 647;
Best Local Similarity 26.7%; Pred.No.0.00084;
Matches 72; Conservative 21; Mismatches 113; Indels 64; Gaps 13;

QY 23 GGYRVKARTSYVDVETLFGSPAGTRPPDPDP-PWVEKANRTRGVGEKASKALGAKGSC 81
DB 59 GGYNPPSPSICSTFTTGGGGGYTPTSDTPPSPSSDTSPTPGGCGSSPT----PCD 113
QY 82 TTPSRGS--TPLLTPRKNNKVRTPSHTPSYCDESLFGSRSEGASFGAPRMAGDAKLRAL 140
DB 114 APPSPSSDTSPTTTPGGGGGSP--TPS-----DTPPSPSSD----- 147
QY 141 LWTPEPTD--RGSHSPRPEAP-----LRAIHPAG-----PGKTPGPAADSQKLSM 185
DB 148 --TSPTTGGGGGYTPTSDAPSPSSDTSFTTGGGGGYTPTSDAPSPSSDTSPTTP 205
QY 186 GGLHSSRLPKRLGSLHNLNVSTPGHPATSPHTNGPDLPSTG-----GVTFR 236
DB 206 GGGGGYTPPTSDTPPSPSSGSPPT--PGGGGYTPTSDTPPSPSSGSSRTTPGCGCTP 263
QY 237 SPLVTSRARSVSISVPTPRGGATQPKP 266
DB 264 TPCGTPRAPSSGTS-PTTP--GGSYYPPT 290

RESULT 5
ID Q8MQG8 PRELIMINARY; PRT; 518 AA.
AC Q8MQG8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PQN-75 protein (corresponding sequence W0302.1b).
DE W0302.1 OR PQN-75.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
CX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RC MEDLINE=99069613; PubMed=9851916;
RX Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018 (1998).
```

[2]		SEQUENCE FROM N.A.	
RNP		STRAIN-Bristol N2;	
RC		Rohlfing T., Wohldmann P.;	
RA		"The sequence of C. elegans cosmid W03D2.";	
RL		Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.	
[3]			
RNP		SEQUENCE FROM N.A.	
RC		STRAIN-Bristol N2;	
RA		Waterston R.;	
RL		Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.	
RNP		EMBL; AF000298; AAM97960.1; -	
RC		WormPep; W03D2.1b; CE31729.	
RA		SEQUENCE 518 AA; 50473 MW; 4801F27D9663651E CRC64;	
RL			
SQ			
Query Match		10.1%; Score 144.5; DB 5; Length 518;	
Best Local Similarity		24.1%; Pred. No. 0.0018;	
Matches 75; Conservative 18; Mismatches 97; Indels 121; Gaps 13;			
OY	13	QTLLGLHCRGCGVVRKARTSYVDTELFSGPACTRTPTDFDPWPVEKANRTRGVGKEASK 72	:
Ddb	213	QGFGQQCQTONGFRGRKQ-----APPAGSPPPP--PP-----XGSP 247	:
OY	73	ALGAKSCETTPRGSGTPTLTRKNKYRPISHITSYCDLSFGSRSEGSFGAPRNAGK 132	:
Ddb	248	PLAGSGSPPPPPAAGSP--PRTGSPPPPTGSP-----PPPPAGG 287	:
OY	133	DAAKLRALLWTPTPTPRGS-----HSRP-----REAPLRAIHPA 167	:
Ddb	288	SPPPPRAGSPPPPPPRRGSPPTGSLUPPGAAGSPPAGTGSPPPPQRCKQAPEPSFTG 347	:
OY	168	GPKSTEP-----GFAADSOKLS-----MG-----LHSSRLKRL 198	:
Ddb	348	SPPTGSPPTGRPPRGPGFGKSSESSESRERGPRGGPRGPRGXSESSESEPREPRGP 407	:
OY	199	SHSLTHLVNSTGHATSAFHNGPDLPBSTSGVTFSPLVTSARSVISVPSTPRRG 258	:
Ddb	408	RRS-----PPTGSPPTGSPPTGRPPRGSPPTG-----SPPTGLPSRKQRQ 447	:
OY	259	GATQKP--KPP 267	:
Ddb	448	APERDTGSPFP 458	:
RESULT 6			
ID	002123	PRELIMINARY; PRT; 524 AA.	
AC	002123;		
DT	01-JUL-1997 (TrEMBLrel. 04, Created)		
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)		
DE	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Hypothetical protein.		
GN	W03D2.1.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;		
OC	Rhabditiidae; Peioderinae; Caenorhabditis.		
OX	NCBI_Taxid=6239;		
[1]	(1)		
RN	RP	SEQUENCE FROM N.A.	
RC	STRAIN-Bristol N2;		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	None;		
RT	"Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";		
RL	Science 282:2012-2018(1998).		
[2]	(2)		
RN	RP	SEQUENCE FROM N.A.	
RC	STRAIN-Bristol N2;		
RA	Rohlfing T., Wohldmann P.;		
RL	"The sequence of C. elegans cosmid W03D2.";		
RT	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.		
[3]	(3)		
RN	RP	SEQUENCE FROM N.A.	

RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission."; to the EMBL/GenBank/DBJ databases.  
 RL Submitted (S2P-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF000298; AAC48255.2; -;  
 DR WormPep; W03D2.1a; CE28501.  
 KW Hypothetical protein.  
 SQ SEQUENCE 524 AA; 51210 MW; 023D77EFPD172CC5 CRC64;

Query Match 10.1%; Score 144.5; DB 5; Length 524;  
 Best Local Similarity 24.1%; Pred. No. 0.0018;  
 Matches 75; Conservative 18; Mismatches 97; Indels 121; Gaps 13;

QY 13 QTLGLQHCRCGYRVKARTSYVDETFLFGSPAGTPTPTDPPWVEKANRTRGVGEASK 72  
 Db 219 QGFGQQQTQNGFRGKRQ-----APPAGSPPPP-----PP-----KGSP 253  
 QY 73 ALGAKGSCETTPSRGSTTTLTPRKKNKYRPISTHTPSYCDLSFGSRSEGASFGAPRMAG 132  
 Db 254 PLAGSGSPPPPAAAGSP--PPRTGSPPPPTGSP-----PPPPAGG 293  
 QY 133 DAAKLALLWTPPTPRGS-----HSPRP-----REAPLRAIHPA 167  
 Db 294 SPPPRAGSPPPPPPPKSGSPPTGLPPQAGGSPPPAGTSPPPPRQKQAPERSPPTG 353  
 QY 168 GPSKTEP-----GPAADSQKLS-----MG-----LHSSRPLKRG 198  
 Db 354 SPTGSPPTGPRGPGKSGSE 413  
 QY 199 SHSLTHLNVSTGHPATSAHTNGPDLRSTSGVTFRSPLVTSRARSVISVPSTPRRG 258  
 Db 414 RRS-----PPTGSPPTGSPPTGPRGSPPTG-----SPTGLPFRQKRQ 453  
 QY 259 GATOKP--KPP 267  
 Db 454 APEDRPTGSP 464

## RESULT 7

Q8MQG9  
 ID Q8MQG9 PRELIMINARY; PRT; 539 AA.  
 AC Q8MQG9;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE PQN-75 protein (corresponding sequence W03D2.1c).  
 GN W03D2.1 OR PQN-75.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA Waterston R.;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RC Rolfing T.; Wohldmann P.;  
 RT "The sequence of C. elegans cosmid W03D2.";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF000298; AAM97961.1; -;  
 DR WormPep; W03D2.1c; CE14506.

SQ SEQUENCE 539 AA; 52555 MW; D3D3CFD8EF42CDB3 CRC64;

Query Match 10.1%; Score 144.5; DB 5; Length 539;  
 Best Local Similarity 24.1%; Pred. No. 0.0019;  
 Matches 75; Conservative 18; Mismatches 97; Indels 121; Gaps 13;

QY 13 QTLGLQHCRCGYRVKARTSYVDETFLFGSPAGTPTPTDPPWVEKANRTRGVGEASK 72  
 Db 224 QGFGQQQTQNGFRGKRQ-----APPAGSPPPP-----PP-----KGSP 268  
 QY 73 ALGAKGSCETTPSRGSTTTLTPRKKNKYRPISTHTPSYCDLSFGSRSEGASFGAPRMAG 132  
 Db 269 PLAGSGSPPPPAAAGSP--PPRTGSPPPPTGSP-----PPPPAGG 308  
 QY 133 DAAKLALLWTPPTPRGS-----HSPRP-----REAPLRAIHPA 167  
 Db 309 SPPPRAGSPPPPPPPKSGSPPTGLPPQAGGSPPPAGTSPPPPRQKQAPERSPPTG 368  
 QY 168 GPSKTEP-----GPAADSQKLS-----MG-----LHSSRPLKRG 198  
 Db 369 SPTGSPPTGPRGPGKSGSE 428  
 QY 199 SHSLTHLNVSTGHPATSAHTNGPDLRSTSGVTFRSPLVTSRARSVISVPSTPRRG 258  
 Db 429 RRS-----PPTGSPPTGSPPTGPRGSPPTG-----SPTGLPFRQKRQ 468  
 QY 259 GATOKP--KPP 267  
 Db 469 APEDRPTGSP 479

## RESULT 8

Q26963  
 ID Q26963 PRELIMINARY; PRT; 964 AA.  
 AC Q26963;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Trans-sialidase.  
 GN TCTS-121.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=Y;  
 RA Uemura H.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 178-378 FROM N.A.  
 RC STRAIN=Y;  
 RX MEDLINE=93010978; PubMed=1396577;  
 RA Uemura H.; Schenkman S.; Nussenzweig V.; Eichinger D.;  
 RT "Only some members of a gene family in Trypanosoma cruzi encode  
 RT proteins that express both trans-sialidase and neuraminidase  
 RT activities.";  
 RL EMBO J. 11:3837-3844 (1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=Y;  
 RC MEDLINE=97001677; PubMed=8844669;  
 RA Smith L.E.; Uemura H.; Eichinger D.;  
 RT "Isolation and expression of an open reading frame encoding sialidase  
 RT from Trypanosoma rangeli.";  
 RL EMBL; D50684; BAA09333.1; -;  
 DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.  
 DR GO; GO:0009405; P:pathogenesis; IEA.  
 DR InterPro; IPR008985; CoA-like\_lec\_gl.  
 DR InterPro; IPR002860; GH\_BNR.  
 DR InterPro; IPR008377; Sialidase\_trypan.  
 DR Pfam; PF02012; BNR; 3.  
 DR PRINTS; PR01803; TCSIALIDASE.



SQ SEQUENCE 964 AA; 101611 MW; 8F0600289FB847AD CRC64;

Query Match 9.9%; Score 142.5; DB 5; Length 964;  
 Best Local Similarity 26.8%; Pred. No. 0.0053;  
 Matches 84; Conservative 24; Mismatches 109; Indels 97; Gaps 19;

QY 13 QTLGLQRCGGYR-----VKARTSYVDLTFGSPA-----43

DB 609 RTPDISHYVGGYRSDMPISHTVNNVLLYNQLNAEIRLFLSQDLIGTEAHWDSS 668

QY 44 -----GTRTPPPDFPPWVEKANRTRGV-----GKEASKALGAKGSCETTPS-----RG 87

DB 669 SDSNAHSTPSTPGD-----SSAHSSTPSTPADNGAHSTPSTPADNGAHSTPSTPGDNGAH 722

QY 88 STPTLTTRKKNKYPIPIHTSYCDSESLFGSRSEGAISGAPRMAGDAKLALLWTPPPT 147

DB 723 STP-LTPADNGAH-----STPTTPADSSAHSSTPSTPADNGA-----HSTPST 763

QY 148 P-RGSHSPRPREAPLRAIHPA--GPSKTEPGPAADSQKLSM-----GGLHS--GRPLK 195

DB 764 PADNGAHS-----TPLT---PADNGAHSTPPT-ADSSAHSSTPSTPADNGAHSTPSTPAD 814

QY 196 RGLSHSLTHLNVPTGHPATGAPTNGQD--LRPSTSGVTFRPLVTSRARSVSISVPS 253

DB 815 NG-AHSTPPTPADSSAHSSTPSTPGDNGAHSTPSTPADNG-AHSTPL--TPADSSAHSSTPS 870

QY 254 TPRGGATQKPKPP 267

DB 871 TPGDNGAHSTPSAP 884

RESULT 9

Q9LLJ1 PRELIMINARY; PRT; 769 AA.

AC Q9LLJ1

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE CALK protein.

OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

OC Chlamydomonadaceae; Chlamydomonas.

OX NCBI\_TaxID=3055;

RN NCBI\_TaxID=3055;

RA Pan J., Snell W.J.;

RP MEDLINE=20379056; PubMed=10807915;

RT "Regulated targeting of a protein kinase into an intact flagellum. An aurelia/iplp-like protein kinase translocates from the cell body into the flagella during gamete activation in Chlamydomonas."

RL J. Biol. Chem. 275:24106-24114(2000).

CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; AF199021; AAF97501.1; -.

DR HSSP; Q63450; 1A06.

DR GO; GO:0005524; P:ATP binding; IEA.

DR GO; GO:0004674; P:protein serine/threonine kinase activity; IEA.

DR GO; GO:0016740; P:transferase activity; IEA.

DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.

DR InterPro; IPR000719; Ser thr kinase.

DR InterPro; IPR002290; Ser thr kinase.

DR InterPro; IPR008271; Ser\_thr\_kin\_AS.

DR Pfam; PF00069; pkinase\_1.

DR Prodom; PD000001; Prot\_kinase; 1.

DR SMART; SM00220; S\_TKc; 1.

DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.

DR PROSITE; PS00108; PROTEIN KINASE ST; 1.

KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

SQ SEQUENCE 769 AA; 80250 MW; DB037FFCD4245101 CRC64;

Query Match 9.5%; Score 136; DB 10; Length 769;  
 Best Local Similarity 22.8%; Pred. No. 0.013;  
 Matches 75; Conservative 41; Mismatches 115; Indels 98; Gaps 15;

QY 32 SYVDETLFG-----SPAGTRTPPPDFDP-----WVEKANR-----62

DB 323 SEVEPRQLGFRGSNAPGAGEPSGPKSEPPRPLQVSSVAQGGFLCHSSMSAMDCWRYD 382

QY 63 -----TRGVGKEASKAL-GAKGSCETTPSRGSTPTLTTRKKNK-----VRPISH 105

DB 383 IHLDSVTGTMASAAATEALKSVKRVSDGDTTGGPRSAAGGRVSRAGTARNSTGQAPLSR 442

QY 106 TSYCDESLFGSRSE-----GASFGAPRMAGDAKLALLW-----TPPTPRGS--- 151

DB 443 MNMSAQAVAAAAAAMASGSGFGAAGHSEDSPKGRALTREELYASITAPLGGSGPVL 502

QY 152 -----HSPRPRE-AFLRAIHPAGPKTEPGPAADSQKLSMGLHSRPLKRG-----LS 199

DB 503 ARSPTAGRASPAPOQWAPL-----PTNACKKSGGGGLSGRSLSQATCATGAPPSRGAELIA 558

QY 200 HSLTHLNVPTGHP-----PATSA-----PHTNG-----PODLRPTSTSGVTFRSP 238

DB 559 HDLSRLRLRVDASAASGGGGAAGPSTSGAVPYSPGKSSGIRPSPPOAPSPRTARGLTGSP 618

QY 239 LVTSRARSVSISVPSTPRRGGATQKPKPP 267

DB 619 LTPSL-SQSFTSPVSPPLAAGSNASATP 646

RESULT 10

Q9VLW2 PRELIMINARY; PRT; 1131 AA.

AC Q9VLW2

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE CG14535 protein.

DE CG14535.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN NCBI\_TaxID=7227;

RP SEQUENCE FROM N.A.

RC STRAIN=Berkely;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galte R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt J.C., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Chery J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Sidon-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Gelniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banton J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferrieria S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Williams S.M., Strong R., Svirkas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Kamin R., Carlson J.W., Celnikier S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Mallburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Snu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003619; AAF52589.2;  
DR FlyBase; FBgn0031955; CG14535.  
DR GO; GO:0005871; C:kinesin complex; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0003774; F:motor activity; IEA.  
DR GO; GO:0007017; P:microtubule-based process; IEA.  
DR InterPro; IPR001752; P:kinasin\_motor.  
DR Pfam; PF00225; kinesin; 1.  
DR SMART; SM00129; KISC; 1.  
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.  
SQ SEQUENCE 1131 AA; 120721 MW; CB8413CD0F79466C CRC64;  
Query Match  
Best Local Similarity 9.3%; Score 134; DB 5; Length 1131;  
Matches 70; Conservative 30; Mismatches 99; Indels 56; Gaps 13;  
QY 2 KTPVELAVSGNQTGLQHCRCG-GYRVKARTSVYDETLFGSP---AGTRTPDPDPWV 57  
DB 402 RVPMFLAVGLAQGLGNGSSAGSDPSSEISADTVIYMGPNDDATDGEHPVLPST 461  
QY 58 EKANRTGVCKEASKALGKGS-CEITPSR--GSTPTLTPR----KKVVRPISHTPSYC 110  
DB 462 --AGDNKGV---MSKAL--KSGLEKTPSKASNSPMKKMAAKAKKLPCHSGTSL- 513  
QY 111 DESLFGSRSGASGAPRMAKGAALALLWTTPPTPR-----GSHSRPRPREAPLRAH 165  
DB 514 -----KQAGAGACSSPMIPHEQPQIQ-A-MGSPFIPRHMVSKGMSFSPKSPMRAH 566  
QY 166 P-----AGPSKTEPGCPAADSKQLSMGGLHSRPLKGLSH----- 200

DB 567 PGALBQLEAGNKKITEEOWIDGPRVSRKAVAEARHLMREVNHVHVKQETWVDGPKSQSCR 626  
QY 201 SLTHLNVPTGHPAT 215  
DB 627 SLTACNLPAAGGSQT 641  
RESULT 11  
Q7SY42 PRELIMINARY; PRT; 1303 AA.  
AC Q7SY42;  
DT 01-OCT-2003 (TREMBLrel. 25, Created)  
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypothetical protein  
OS Brachydanio rerio (zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AB; TISSUE=Body;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Wang J., Heide F.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.J., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., Morley P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AB; TISSUE=Body;  
RA Strausberg R.;  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC055134; AAH55134.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 1303 AA; 147812 MW; F10EF49F81EA7E17 CRC64;  
Query Match  
Best Local Similarity 9.3%; Score 133.5; DB 13; Length 1303;  
Matches 58; Conservative 31; Mismatches 110; Indels 41; Gaps 8;  
QY 50 PDFPPVWEKANRTGVCKEASKALGKGSCEITPSRGSTPTLTPRKN----- 98  
DB 498 PSDKPAWAKVEERSKLNQSSPALQHK-----VANRISDFSLPPRSSESGMQPART 552  
QY 99 --KYRISHTPSYCDSEIFGSRSESGASFGA-----PRWAKGDAKRALLTPTPTPRGS 151  
DB 553 PMERPVEPQMAH-----LVFVRSHSSMSVSQSLHDSAGKVSFAQESLVSNERPQRN 608  
QY 152 HSP---RPREAPLRAHPAGSKTEPGAADSKQLSMGGLHSRPLKGLSHSLTHLNV 208  
DB 609 SDPTSEMPPPAPRPMASHEEKPDSSPWLREDAVPPKLSGTTQVFPQRTTSIPALARKNSP 668  
QY 209 ST-----GHPATSAPTNGPDQLRPSTGVTPRSLVTSRARSVTSVSTPRRGATQ 262  
DB 669 NGSSSGGGRASQLIRTSNP-DLURTELSL-----DVALQRTSSNSSSSSTSSQGGGQ 723

```

RESULT 12
Q9JIK1 PRELIMINARY; PRT; 862 AA.
AC Q9JIK1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 25, Last annotation update)
DE Mu-protocadherin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=20387356; PubMed=10801787;
RA Goldberg M., Peshkovsky C., Shifteh A., Al-Awqati Q.;
RT "mu-protocadherin, a novel developmentally regulated protocadherin
RT with mucin-like domains.";
RL J. Biol. Chem. 275:24622-24629 (2000).
DR EMBL; AF221952; AAF70456.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR PRINTS; PRO0205; CADHERIN.
DR SMART; SM00112; CA; 3.
DR PROSITE; PS00232; CADHERIN_1; 1.
DR PROSITE; PS0268; CADHERIN_2; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
SQ SEQUENCE 862 AA; 90376 MW; 302E302503A9C375 CRC64;

Query Match 9.3%; Score 133; DB 11; Length 862;
Best Local Similarity 27.7%; Pred. No. 0.027;
Matches 66; Conservative 22; Mismatches 82; Indels 68; Gaps 14;

QY 42 PAGTRPTPPDPDPWVEKANTRGVGKEASKALGAKGSCETTPSRGSTPTLTTPRKKNYR 101
DB 460 PSTESPTTPEAGGTGSSNTT-----LETSTSTSGSPATSSGGAGFPPP-AGTTLS 514
QY 102 PISHTPSYCDLSFGSRSEGFAGPRMAK--GDAKLRALLMTTP-----PTPRGS 151
DB 515 PLTSAPTVPG---GSPTLGIS-TSPQTATPGGATQ-----TPKPGTSQPMVPTPGAS 563
QY 152 HSPRPRAPLRAHPAGSKTEPGPAADSKLSMGLGHSRPLKGLSHSLTHLNVPTSG 211
DB 564 TSSQFAT-----PSGSSTQTPKPG-----TSQPM-----VPTPG 592
QY 212 HPATSAPHT---NGPQDLRPSTSGVTPRSPLVTSRASVSISVPSPTPRRGATQKPKP 266
DB 593 ASTSSQATPSGSSTQTPRPTS-----QPMVPTPGASTS-SQATP--SGSTQTPKP 642

RESULT 13
Q9DWH8 PRELIMINARY; PRT; 1240 AA.
AC Q9DWH8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE R2.
GN R2.
OS Rat cytomegalovirus (strain Maastricht).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Nuromegalovirus.
OX NCBI_TaxID=79700;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Maastricht;
RX MEDLINE=20366325; PubMed=10906222;
RA Vink C., Beuken E., Bruggeman C.A.;
RT "Complete DNA sequence of the rat cytomegalovirus genome.";

J. Virol. 74:7656-7665 (2000).
[2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Maastricht;
RX MEDLINE=20473137; PubMed=11018281;
RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A., Vink C.;
RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a
RT spliced transcript.";
RL Virus Res. 69:119-130 (2000).
DR EMBL; AF232689; AAF99111.1; -.
DR EMBL; AF232689; AAF99111.1; -.
SQ SEQUENCE 1240 AA; 125612 MW; 33B6C13DC6A272B0 CRC64;

Query Match 9.3%; Score 133; DB 12; Length 1240;
Best Local Similarity 24.0%; Pred. No. 0.04;
Matches 74; Conservative 19; Mismatches 115; Indels 100; Gaps 13;

QY 42 PAGTRPTPPDPDPWVEKANTRGVGKEASKALGAKGSCETTPSRGSTPTLTTPRKKNYR 101
DB 250 PAAPPPPPQTPQPSVAPPHRRFSSSTLASLTFDGA-GVGVGLGPTTTPAGTTPVYQ 308
QY 102 PISHTPSYCDLSFGSRSEGFAGPRMAKGDAA-----KLRALL 141
DB 309 QPPSTAIY-----GAAAGVFGFADGTAGGAEDDFGDEADLEGRRVFDOLAVVV 357
QY 142 WT----BPPTPRGSHSRPREAPL-----RAHPAG-----PSKTEPGPAA 178
DB 358 RSDDLVPNNHPSGLDVLPAAPATDLRRPQVSPASRGDSLSRGAAGTWDPSVPEAPAG 417
QY 179 DSQ-----KLSMGLHSR-----PLKRG-----LSHS 201
DB 418 PRAAGWPGTSSRG--SSRSCPGRLSDATPATPPASRGDASTEGTTTTTTTTTTTTTT 475
QY 202 LTHLNVPTSGHPATSAHPHTNGPDLPSTSGVTPRSPLVTSRASVSISVPSPT-PRRGA 260
DB 476 TTGLGKGKGTGVLTP-PRTS-----APSTTATSAPTLPSTARSRCDSGPTFGSPRAR 530
QY 261 TQPKPPW 268
DB 531 TSTPDPW 538

RESULT 14
Q87W07 PRELIMINARY; PRT; 488 AA.
AC Q87W07;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type III effector Hoppmai (Pto).
GN HOPPMAI (PTO) OR PSPTO4776.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengay J., Nelson W., Davidse T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016873; AAO58206.1; -.
DR TIGR; PSPT04776; -.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR PROSITE; PS0076; DnaJ_2; 1.
DR Complete proteome.
KW SEQUENCE 488 AA; 52549 MW; 4D402A03A3696DDF CRC64;

Query Match 9.2%; Score 132.5; DB 16; Length 488;
```



Db 241 GGATCCCGATGCTACGAGCCAAAGATGCTCAGCTTTATAGGTGTGACCTACACATGTGAC 300  
Qy 491 TTCACTCAGTTTGTGATCCGTAATAATGGAATAATTCGAAGCTACTTTCACAGTGTGTT 550  
Db 301 TTCACCTCAGTTTGTGATCCGTAATAATGGAATAATTCGAAGCTACTTTCACAGTGTGTT 360  
Qy 551 GAGAGGATTAATGAACAATGCTGTGTAAGCTCTTTGAGAGGAGGAGCTCGGAAGCAG 610  
Db 361 GAGAGGATTAATGAACAATGCTGTGTAAGCTCTTTGAGAGGAGGAGCTCGGAAGCAG 420  
Qy 611 GGCTGCGCGCAGAGACACCTGCTGTCTACCAAGGACACACAGGAGAGCTGAAGACCC 670  
Db 421 GGCTGCGCGCAGAGACACCTGCTGTCTACCAAGGACACACAGGAGAGCTGAAGACCC 480  
Qy 671 GTGAGCTGCGCTCAGTGGGATGCGAGCCTCGGCTTCAGACCGCTCGGAGGTGCG 730  
Db 481 GTGAGCTGCGCTCAGTGGGATGCGAGCCTCGGCTTCAGACCGCTCGGAGGTGCG 540  
Qy 731 TACCGGCTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTTGGAGCCCGCAGGCG 790  
Db 541 TACCGGCTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTTGGAGCCCGCAGGCG 600  
Qy 791 ACCGGCTACCCACCGGACTTCGATCCGCTCGGCTGGTGGAGAGGCTAACGAGACCA 850  
Db 601 ACCGGCTACCCACCGGACTTCGATCCGCTCGGCTGGTGGAGAGGCTAACGAGACCA 660  
Qy 851 GGCTGGGCAAGGAGGATCGAAGGCTTGGGGCAAGGGAGCTGTGAGACCAACCC 910  
Db 661 GGCTGGGCAAGGAGGATCGAAGGCTTGGGGCAAGGGAGCTGTGAGACCAACCC 720  
Qy 911 TCAAGGGGAGGACCCCGCTCACACCAAGGAAGAAACAAATACAGACCCATCAG 970  
Db 721 TCAAGGGGAGGACCCCGCTCACACCAAGGAAGAAACAAATACAGACCCATCAG 780  
Qy 971 CACACCGCTTACTGTGATGATGCTGCTTGGCTCCGATCTCAAGGCGCAGCTTC 1030  
Db 781 CACACCGCTTACTGTGATGATGCTGCTTGGCTCCGATCTCAAGGCGCAGCTTC 840  
Qy 1031 GGGGCCCGCGGATGGCAAGGGGATGCGCGAAAGCTCCGTCCTCTTTGTGGAGCA 1090  
Db 841 GGGGCCCGCGGATGGCAAGGGGATGCGCGAAAGCTCCGTCCTCTTTGTGGAGCA 900  
Qy 1091 CCACCTACCCCGAGGATGACCTGCGCCCGCCCGGAGGAGGACACCTGCGAGCATT 1150  
Db 901 CCACCTACCCCGAGGATGACCTGCGCCCGCCCGGAGGAGGACACCTGCGAGCATT 960  
Qy 1151 CACCCAGCTGCTCCCTCCAGACAGAGCCGGGCGCAGCGCAGACTCCAGAGATTATCT 1210  
Db 961 CACCCAGCTGCTCCCTCCAGACAGAGCCGGGCGCAGCGCAGACTCCAGAGATTATCT 1020  
Qy 1211 ATGGTGGGTATACCTTTCAGGCGCTGAAAGGGGAGCTTCCCATTCCTCACCAC 1270  
Db 1021 ATGGTGGGTATACCTTTCAGGCGCTGAAAGGGGAGCTTCCCATTCCTCACCAC 1080  
Qy 1271 CTGAATGTCGCCAGCTGTGTCATCCAGCCACAGCTGCCCCCACCACAAATGGGCTCAG 1330  
Db 1081 CTGAATGTCGCCAGCTGTGTCATCCAGCCACAGCTGCCCCCACCACAAATGGGCTCAG 1140  
Qy 1331 GATCTCAGGCTTCCAGCTCAGGAGTGAACCTTCGAGAGCCCTGCTGATCTCCAGGCT 1390  
Db 1141 GATCTCAGGCTTCCAGCTCAGGAGTGAACCTTCGAGAGCCCTGCTGATCTCCAGGCT 1200  
Qy 1391 CGCTCAGTTAGCATTTTCAGTGCCATCTACCCACAGAGGTTGGGCGCCACCGAGAACCA 1450  
Db 1201 CGCTCAGTTAGCATTTTCAGTGCCATCTACCCACAGAGGTTGGGCGCCACCGAGAACCA 1260  
Qy 1451 AAGCCCTTTGGAATGATACTCTTTTCATCAGGTTTGCCTATGGGCGCAGGCGACAGT 1510  
Db 1261 AAGCCCTTTGGAATGATACTCTTTTCATCAGGTTTGCCTATGGGCGCAGGCGACAGT 1320  
Qy 1511 ATGGCCCTTGGCAGGTTAGGAGCATTCATCACCCAGGAGACCCAGGATTAAAGAA 1570  
Db 1321 ATGGCCCTTGGCAGGTTAGGAGCATTCATCACCCAGGAGACCCAGGATTAAAGAA 1380

Query Match 88.2%; Score 1597.6; DB 9; Length 1709;  
Best Local Similarity 94.5%; Pred. No. 0;  
Matches 1705; Conservative 0; Mismatches 4; Indels 95; Gaps 1;

ORIGIN

FEATURES

Location/Qualifiers  
1..1709  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMR322005520"  
/cell\_line="IMR32"  
/cell\_type="neuroblastoma"  
/clone\_lib="IMR322"  
/note="cloning vector: pME18SFL3"

COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

TITLE

Unpublished  
2 (bases 1 to 1709)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

JOURNAL

Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

REFERENCE

1  
Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEO human cDNA sequencing project

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

KEYWORDS

oligo capping; fis (full insert sequence).

VERSION

AK123762.1 GI:34529382

DEFINITION

Homo sapiens cDNA FLJ41768 fis, clone IMR322006520.

LOCUS

AK123762  
1709 bp mRNA linear PRI 09-SEP-2003

Qy

1571 GCCCTGTGGGGGAGAGACATAGCAGGGTGGGAGTGCCTCCCTTTATCCTGACAA 1630

Db

1381 GCCCTGTGGGGGAGAGACATAGCAGGGTGGGAGTGCCTCCCTTTATCCTGACAA 1440

Qy

1631 TCTCTAGTCGATTTCTGCTTTTCTCCGATTCGGGATTTGGGGGCGACCTTAAGATG 1690

Db

1441 TCTCTAGTCGATTTCTGCTTTTCTCCGATTCGGGATTTGGGGGCGACCTTAAGATG 1500

Qy

1691 CCTCTCTCAGCCCTGTCTCAACCATCTCAAAATTAGTGCACACCCAGGGCGCTGCAC 1750

Db

1501 CCTCTCTCAGCCCTGTCTCAACCATCTCAAAATTAGTGCACACCCAGGGCGCTGCAC 1560

Qy

1751 CTCCACATCATTCATTCCTTGTCTGTCGCAAGTGGGAATAAAGCGCGTATTGCCAACCTG 1810

Db

1561 CTCCACATCATTCATTCCTTGTCTGTCGCAAGTGGGAATAAAGCGCGTATTGCCAACCTG 1620

Qy

1811 G 1811

Db

1621 G 1621

QY 851 GGCCTGGGCAAGAGGATCGAAGCCCTTGGGGCAAGAGGAGCTGTGAGACACACCCC 910  
 Db 661 GGCCTGGGCAAGAGGATCGAAGCCCTTGGGGCAAGAGGAGCTGTGAGACACACCCC 720  
 QY 911 TCAAGGGCAGACACCCACCCTCACCAAGAGAGAGAACAAATACAGACCCATCAGC 970  
 Db 721 TCAAGGGCAGACACCCACCCTCACCAAGAGAGAGAACAAATACAGACCCATCAGC 780  
 QY 971 CACACCCCGTCTTACTGTGATGAGTCGTCTTCCGCTCCCGATCTGAAGCGCCAGCTTC 1030  
 Db 781 CACACCCCGTCTTACTGTGATGAGTCGTCTTCCGCTCCCGATCTGAAGCGCCAGCTTC 840  
 QY 1031 GGGGCCCCCGGATGGCGAAGGGGATGCCCAAGCTCGTGCTCTTGTGTGACCCCA 1090  
 Db 841 GGGGCCCCCGGATGGCGAAGGGGATGCCCAAGCTCGTGCTCTTGTGTGACCCCA 900  
 QY 1091 CCACCTACCCCGGATGAGCACTCGCCCGCCCGCCAGGAGGCCACCACTCGGAGCCATT 1150  
 Db 901 CCACCTACCCCGGATGAGCACTCGCCCGCCCGCCAGGAGGCCACCACTCGGAGCCATT 960  
 QY 1151 CACCAGCTGGTCCCTCCAGACAGAGCGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1210  
 Db 961 CACCAGCTGGTCCCTCCAGACAGAGCGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1020  
 QY 1211 ATGGGTGGGTACACTCTTCAAGCCCTTGAAGCGGGACTTTCCCAATTCCTCACCAC 1270  
 Db 1021 ATGGGTGGGTACACTCTTCAAGCCCTTGAAGCGGGACTTTCCCAATTCCTCACCAC 1080  
 QY 1271 CTGAATGCCCGACACTGTCATCCAGCAGGAGTCCCGCCACACAAATGGGCTTCAG 1330  
 Db 1081 CTGAATGCCCGACACTGTCATCCAGCAGGAGTCCCGCCACACAAATGGGCTTCAG 1140  
 QY 1331 GATCTAGGCTTCCAGTCAAGGGTGAATCTTCCGAGGAGGAGGAGGAGGAGGAGGAGG 1390  
 Db 1141 GATCTAGGCTTCCAGTCAAGGGTGAATCTTCCGAGGAGGAGGAGGAGGAGGAGGAGG 1200  
 QY 1391 CGCTCAGTACATTTAGTGGCATCTACCCAGCAGAGGTGGGGCACCAGAACCA 1450  
 Db 1201 CGCTCAGTACATTTAGTGGCATCTACCCAGCAGAGGTGGGGCACCAGAACCA 1260  
 QY 1451 AAGCCCCCTTGAATGATCTCTTTTCATCAGGCTTGCCTATGGGGCCAGCGGACAGGT 1510  
 Db 1261 AAGCCCCCTTGAATGATCTCTTTTCATCAGGCTTGCCTATGGGGCCAGCGGACAGGT 1320  
 QY 1511 ATGGCCCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1570  
 Db 1321 ATGGCCCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380  
 QY 1571 GCCCTGTGGGGCAGACAGACATAGAGGGGTGGGAGTGCCTCTTATCTGACAA 1630  
 Db 1381 GCCCTGTGGGGCAGACAGACATAGAGGGGTGGGAGTGCCTCTTATCTGACAA 1440  
 QY 1631 TCTCTAGTCGATTTCTGCTTTTCTCCGATTCGAGATTTGGGGCCACTCTTAAGATG 1690  
 Db 1441 TCTCTAGTCGATTTCTGCTTTTCTCCGATTCGAGATTTGGGGCCACTCTTAAGATG 1500  
 QY 1691 CCTCTCTCCAGGCTGTCTCAACCATCTCAATTTAGTCCCAACCCAGGGGCTGGCAC 1750  
 Db 1501 CCTCTCTCCAGGCTGTCTCAACCATCTCAATTTAGTCCCAACCCAGGGGCTGGCAC 1560  
 QY 1751 CTCACCATCATCTGCTTGTGTGCGAAGTGGGAATAAACCGCGGTGATTGCGCAACTG 1810  
 Db 1561 CTCACCATCATCTGCTTGTGTGCGAAGTGGGAATAAACCGCGGTGATTGCGCAACTG 1620  
 QY 1811 G 1811  
 Db 1621 G 1621

RESULT 5

AK027733

LOCUS

DEFINITION

Homo sapiens

cDNA

FLJ14827

fis,

clone

OVARC1000886.

PRI

01-AUG-2002

## ACCESSION

AK027733.1

VERSION

GI:14042629

KEYWORDS

oligo capping; fis (full insert sequence).

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1

AUTHORS

Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Saito, Y., Kodaira, H., Kondo, H., Wagatsuma, M., Takahashi, N., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.

TITLE

NEDO human cDNA sequencing project

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 1775)

AUTHORS

Isogai, T. and Otsuki, T.

TITLE

Direct Submission

JOURNAL

Submitted (10-MAY-2001)

COMMENT

Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- &amp; 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES

Location/Qualifiers

1..1775

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="OVARC1000886"

/tissue\_type="ovary, tumor tissue"

/clone\_lib="OVARC1"

/note="cloning vector: pME18SFL3"

469..1278

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BAB55328.1"

/db\_xref="GI:14042630"

/translation="MKTFVELAVSGMOTLGLQHRRCGRYRVKARTSYVDETLFGSPAG TRTPDPDPPEVKANTRGVGEAKGALGAKGSCETTPSRGSTPLTPPKKKYRP ISHTPSYDESLFGSRSEGAPRMAKDAKLRLMTPTTPPRGSHSPREAP LRAIHPGSKTEPGPAADSKLMSGLHSSRLKRLGSLTLHLNVPSTGHSPATAP HTNGPQDLRPSTSGVTFRSLVTSRARSVSISVPSTPRGGATQKPPPWK"

CDS

ORIGIN

Query Match

Best Local Similarity

Matches 1621; Conservative

89.5%; Score 1621; DB 9; Length 1775;

Pred. No. 0;

Mismatches 0;

Indels 0;

Gaps 0;

QY 191 GTGCAAGTCTGGGTTCTGGGTTTCTGGATTTCGGGGCCGCTTACAGTACCTGTGCC 250

Db 1 GTGCAAGTCTGGGTTCTGGGTTTCTGGATTTCGGGGCCGCTTACAGTACCTGTGCC 60

QY 251 GGCTCTCTGGGTGAGTCGTCGCGCGGTGCCCGGGACGGCTTAGGCTGCCGGGGGT 310

Db 61 GGCTCTCTGGGTGAGTCGTCGCGCGGTGCCCGGGACGGCTTAGGCTGCCGGGGGT 120

QY 311 CCGGGCCCCCAGGCATTCGGGCTCAGATTGACGGGGATCCCGATCCACCGCGGCC 370

Db 121 CCGGGCCCCCAGGCATTCGGGCTCAGATTGACGGGGATCCCGATCCACCGCGGCC 180

QY 371 CCGCGCCCTCACCGACGGGTCCAGACCTGGTGGGAAGAGTGGGGACCGGTCCTTGA 430

Db 181 CCGCGCCCTCACCGACGGGTCCAGACCTGGTGGGAAGAGTGGGGACCGGTCCTTGA 240

QY 431 GGATCCGATGCTACGAGCCAGATGCTCAGCTTATAGTGTGACCTACATGTCAC 490

Db	781	CACACCCCGCTTACTGTGATGATCGCTGTTTGGCTCCGATCTGAAGCGCCAGCTTC	840
Qy	1031	GGGGCCCCGGGATGCGAAGGGGATGCGCAAGCTCCGTGCTCTTTGTGGACGCA	1090
Db	841	GGGGCCCCGGGATGCGAAGGGGATGCGCAAGCTCCGTGCTCTTTGTGGACGCA	900
Qy	1091	CCACCTACCCAGGGGTAGCCACTCGCCCGCCCGCCAGAGGACCACTGCGAGCCATT	1150
Db	901	CCACCTACCCAGGGGTAGCCACTCGCCCGCCCGCCAGAGGACCACTGCGAGCCATT	960
Qy	1151	CACCCAGCTGGTCCCTCCAAAGACAGAGCCGGGCGCCAGCGGACACTCCAGAAAGTTATCT	1210
Db	961	CACCCAGCTGGTCCCTCCAAAGACAGAGCCGGGCGCCAGCGGACACTCCAGAAAGTTATCT	1020
Qy	1211	ATGGGTGGGTACACTCTTACGCCCCCTGAAGCGGGACTTCCATTCCTCCACCCAC	1270
Db	1021	ATGGGTGGGTACACTCTTACGCCCCCTGAAGCGGGACTTCCATTCCTCCACCCAC	1080
Qy	1271	CTGAATGTCCCCAGCACTGTCTATCCAGGACACAGTGCCTCCACACAAATGGGCTCAG	1330
Db	1081	CTGAATGTCCCCAGCACTGTCTATCCAGGACACAGTGCCTCCACACAAATGGGCTCAG	1140
Qy	1331	GATCTCAGGCTTCCAGCTCAGGGGTGACTTCCGAGGCCCCCTGTGTGACTTCCAGGGCT	1390
Db	1141	GATCTCAGGCTTCCAGCTCAGGGGTGACTTCCGAGGCCCCCTGTGTGACTTCCAGGGCT	1200
Qy	1391	CGTCTAGTTAGCACTTTCAGTGCCATCTACCCACGACGAGTGGGGCCACCCAGAAACCA	1450
Db	1201	CGTCTAGTTAGCACTTTCAGTGCCATCTACCCACGACGAGTGGGGCCACCCAGAAACCA	1260
Qy	1451	AAGCCCCCTTGGAAATGATACTCTTTTCATCAGGGTTGCCCTATGGGGCCACGCGCAGGT	1510
Db	1261	AAGCCCCCTTGGAAATGATACTCTTTTCATCAGGGTTGCCCTATGGGGCCACGCGCAGGT	1320
Qy	1511	ATGGCCCCCTTGCAGGGTAGGAGCACTTATCACCAGGGGACCCAGGATTAAGAA	1570
Db	1321	ATGGCCCCCTTGCAGGGTAGGAGCACTTATCACCAGGGGACCCAGGATTAAGAA	1380
Qy	1571	GCCCTCTGTGGGGCAGACACATAGCAGGGGTGGGAGTGCCTCCCTTTATCCTGACAA	1630
Db	1381	GCCCTCTGTGGGGCAGACACATAGCAGGGGTGGGAGTGCCTCCCTTTATCCTGACAA	1440
Qy	1631	TCCTAGTCGATTCCTGCTTTTCTCCGATTCGGATTTGGGGCCACCTCTAAGATG	1690
Db	1441	TCCTAGTCGATTCCTGCTTTTCTCCGATTCGGATTTGGGGCCACCTCTAAGATG	1500
Qy	1691	CCTCTCTCCAGCCCTGTCTCAACATCTCCAAATTTAGTGCACACCCAGGGGCTGGCAC	1750
Db	1501	CCTCTCTCCAGCCCTGTCTCAACATCTCCAAATTTAGTGCACACCCAGGGGCTGGCAC	1560
Qy	1751	CTCCACATCATCATGCTTGTGTCGAAGTCGAATAAAGCGCTGATTCGCAACCTG	1810
Db	1561	CTCCACATCATCATGCTTGTGTCGAAGTCGAATAAAGCGCTGATTCGCAACCTG	1620
Qy	1811	G 1811	
Db	1621	G 1621	
RESULT 5	BD158346	1775 bp	DNA linear
LOCUS	BD158346	1775 bp	DNA linear
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.		
ACCESSION	BD158346		
VERSION	BD158346.1	GI:27864104	
KEYWORDS	JP 2002191363-A/13189		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 1775)		
	Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,		
	Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.		

TITLE	Primer for synthesizing full-length cDNA and use thereof
JOURNAL	HELIX RESEARCH INSTITUTE
COMMENT	OS Homo sapiens (human) PN JP 2002191363-A/13189 PD 09-JUL-2002 PF 28-JUL-2000 JP 2000280990 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU, PI KEIICHI NAGAI, TETSUJI OTSUKI PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC 10, C12P21/02, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof PH Key
FT	CDS
source	Location/Qualifiers 1..1775 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
ORIGIN	Query Match 89.5%; Score 1621; DB 6; Length 1775; Best Local Similarity 100.0%; Pred. No. 0; Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0; Qy 191 GTGCAAAAGTGTGGGTTCTGGGTTTCTGGATTTCGGGCGGCTTACACGTAGCTGTGCC 250 Db 1 GTGCAAAAGTGTGGGTTCTGGGTTTCTGGATTTCGGGCGGCTTACACGTAGCTGTGCC 60 Qy 251 GGCTCTCGGTTAGTCCGCGCGGCTGCCCGGGACGGCTAGGCTGCCGGGGT 310 Db 61 GGCTCTCGGTTAGTCCGCGCGGCTGCCCGGGACGGCTAGGCTGCCGGGGT 120 Qy 311 CCGGGGCCAGGCAATTCGGGCTGTCAGATTGACGGGGATCCCGGATGCAACGGCGGCC 370 Db 121 CCGGGGCCAGGCAATTCGGGCTGTCAGATTGACGGGGATCCCGGATGCAACGGCGGCC 180 Qy 371 CCGGGCCCTACCGACGGTCCAGACCTGGTGGGAAGAAGTGGGGGACGGGTCCCTGA 430 Db 181 CCGGGCCCTACCGACGGTCCAGACCTGGTGGGAAGAAGTGGGGGACGGGTCCCTGA 240 Qy 431 GGATCCGATGCTTACGAGCCAGATGCTCAGCTTTTATAGGTGTGACCTACATGTGAC 490 Db 241 GGATCCGATGCTTACGAGCCAGATGCTCAGCTTTTATAGGTGTGACCTACATGTGAC 300 Qy 491 TTCACTCAGTTTGTGATCGTAAATGGACAAATTCGAAGTACTTCAAGTGTGTT 550 Db 301 TTCACTCAGTTTGTGATCGTAAATGGACAAATTCGAAGTACTTCAAGTGTGTT 360 Qy 551 GAGAGGATTAAATGAACAATGCTTTTAAAGCTCTTTGACAGGAGGAGCTTCGGAAGCAG 610 Db 361 GAGAGGATTAAATGAACAATGCTTTTAAAGCTCTTTGACAGGAGGAGCTTCGGAAGCAG 420 Qy 611 GGCTGGCGGCGCAGACACCTCTCTGTCACGGGACCAAGGACAGCTAGAACCCCC 670 Db 421 GGCTGGCGGCGCAGACACCTCTCTGTCACGGGACCAAGGACAGCTAGAACCCCC 480 Qy 671 GTGAGCTGGCGGCTCAGTGGGATCGACACCTCGGCTTCAGCACCCCTCGCAGGTGGC 730 Db 481 GTGAGCTGGCGGCTCAGTGGGATCGACACCTCGGCTTCAGCACCCCTCGCAGGTGGC 540 Qy 731 TACCGGTTCAAGGCCAGAGCGTCAATGTGTGATGAGACTCTGTTTGGCAGCCAGAGGC 790 Db 541 TACCGGTTCAAGGCCAGAGCGTCAATGTGTGATGAGACTCTGTTTGGCAGCCAGAGGC 600 Qy 791 ACCGGCTTACCCACCGGACTTCGATCCGCTGGGTGGAGAGGCTTACAGAACCCAGA 850 Db 601 ACCGGCTTACCCACCGGACTTCGATCCGCTGGGTGGAGAGGCTTACAGAACCCAGA 660

```
QY 1079 TTGTGGAGCGCCACCACTACCCAGGGGTAGCCACTCGCCCGCCCGCCAGGAGGACCA 1138
Db 901 TTGTGGAGCGCCACCACTACCCAGGGGTAGCCACTCGCCCGCCCGCCAGGAGGACCA 960
QY 1139 CTGGAGAGCATTCACCCAGCTGGTCCCTCAAGACAGAGCGGGGCGAGGACATCC 1198
Db 961 CTGGAGAGCATTCACCCAGCTGGTCCCTCAAGACAGAGCGGGGCGAGGACATCC 1020
QY 1199 CAGAGATTATCTATGGTGGGTACACTCTTCAGCCCCCTGAAGCGGGGACATTCCTCAT 1258
Db 1021 CAGAGATTATCTATGGTGGGTACACTCTTCAGCCCCCTGAAGCGGGGACATTCCTCAT 1080
QY 1259 TCCCTCACCCACCTGAATGTCCCGAGCACTGGTCTATCCAGCCACAGTGGCCCCCACCACA 1318
Db 1081 TCCCTCACCCACCTGAATGTCCCGAGCACTGGTCTATCCAGCCACAGTGGCCCCCACCACA 1140
QY 1319 AATGGGCTCAGGATCTCAGGCTTCCAGTCAAGGAGTACCTTCGGAGGCCCCCTGGTG 1378
Db 1141 AATGGGCTCAGGATCTCAGGCTTCCAGTCAAGGAGTACCTTCGGAGGCCCCCTGGTG 1200
QY 1379 ACTTCCAGGGCTCGCTCAGTTAGCATTTCACTGCTATCTACCCACAGCAGGTGGGGCC 1438
Db 1201 ACTTCCAGGGCTCGCTCAGTTAGCATTTCACTGCTATCTACCCACAGCAGGTGGGGCC 1260
QY 1439 ACCCAGAAACCAAGCCCCCTTGGAAATGATCTCTTTTCATCAGGGTGTCTATGGGGCC 1498
Db 1261 ACCCAGAAACCAAGCCCCCTTGGAAATGATCTCTTTTCATCAGGGTGTCTATGGGGCC 1320
QY 1499 ACGGCGACAGTATGGCCCCCTTGCAGGTAGGAGGACATTCATCACCGAGGAAACCCCA 1558
Db 1321 ACGGCGACAGTATGGCCCCCTTGCAGGTAGGAGGACATTCATCACCGAGGAAACCCCA 1380
QY 1559 GGTATTAAAGAGCCCCCTGTGGGGGCGACAGACATAGCAGGGGTGGGCGAGTGCCTCCCT 1618
Db 1381 GGTATTAAAGAGCCCCCTGTGGGGGCGACAGACATAGCAGGGGTGGGCGAGTGCCTCCCT 1440
QY 1619 TTATCTGACAACTCTAGTGCATCTTGCCTTTTCTCCGATTCGGATTTGGGGGCC 1678
Db 1441 TTATCTGACAACTCTAGTGCATCTTGCCTTTTCTCCGATTCGGATTTGGGGGCC 1500
QY 1679 ACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAAAATAGTGCACACCCA 1738
Db 1501 ACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAAAATAGTGCACACCCA 1560
QY 1739 GGGGCTGGACCTCCACATCATCTATGCTTGTGCTGCGAAGTGGAAATAAACGGGGTG 1798
Db 1561 GGGGCTGGACCTCCACATCATCTATGCTTGTGCTGCGAAGTGGAAATAAACGGGGTG 1620
QY 1799 ATTGCCAACCTGG 1811
Db 1621 ATTGCCAACCTGG 1633
```

```
RESULT 4
LOCUS AX880378 1775 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 15283 from Patent EP1074617.
ACCESSION AX880378
VERSION AX880378.1 GI:40035114
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primers for synthesising full-length cDNA and their use
Patent: EP 1074617-A 15283 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..1775
source
```

## CDS

```
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
459..1278
/feature="unnamed protein product"
/codon_start=1
/protein_id="CAE90924.1"
/db_xref="GI:40035115"
/translation="MKTPELVAVSGMOTLGLQHRCGGYRVKARTSYVDETLFGSPAG
TRTPDPDPWVKANETRGVKEASKALGKGCSETTPSRGSPPTLTPRKKNKYRP
ISHTSYCDESLFGSRSEGAFAPRMAKGAALKRALLWTPTTFRGSHRPREAP
LRAHPGSPKTEPGFAADSKLSMGUHSRPLKRGSLHSLTHLVNSTGHPTASAP
HTNGPDLRPSTSGVTPRSPLVTSRARSVSIVPSTPRRGATQKPKPPWK"
ORIGIN
Query Match 89.5%; Score 1621; DB 6; Length 1775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 191 GTGCAAGTCTGGGTCTCTGGTCTTCTGGATTCGCGGGCGGTTCACAGTACGCTGTGCC 250
Db 1 GTGCAAGTCTGGGTCTCTGGTCTTCTGGATTCGCGGGCGGTTCACAGTACGCTGTGCC 60
QY 251 GGTCTCTGGGTGAGTCCGTCCGCGCGGTGCCCCGGAGCGGCTAGGCTGCCGGGGT 310
Db 61 GGTCTCTGGGTGAGTCCGTCCGCGCGGTGCCCCGGAGCGGCTAGGCTGCCGGGGT 120
QY 311 CCGGGGCCCCAGGCATTCGGGCTCAGATTACCGGGATCCCGGATGACCGCGCGGCC 370
Db 121 CCGGGGCCCCAGGCATTCGGGCTCAGATTACCGGGATCCCGGATGACCGCGCGGCC 180
QY 371 CCGGCGCTCACCGAGCGGTCCAGCTGTGGGGAAGAGTGGGGGACGGGTCCCTGA 430
Db 181 CCGGCGCTCACCGAGCGGTCCAGCTGTGGGGAAGAGTGGGGGACGGGTCCCTGA 240
QY 431 GGATCCGATGCTACGAGCAAGATGCTCAGCTTTATAGGTGACCTACACATGTGAC 490
Db 241 GGATCCGATGCTACGAGCCAGATGCTCAGCTTTATAGGTGACCTACACATGTGAC 300
QY 491 TTCACTCAGTTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTTCACAGTCTGT 550
Db 301 TTCACTCAGTTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTTCACAGTCTGT 360
QY 551 GAGAGATTAAATGAACAATGCTTTGTAAGCTCTTCGAGGAGGAGCCCTCGGAGCAG 610
Db 361 GAGAGATTAAATGAACAATGCTTTGTAAGCTCTTCGAGGAGGAGCCCTCGGAGCAG 420
QY 611 GGCCTGGCGGCGAGAGCACACCTGCTGTCACAGGAGCACAGGAGCATGAAGACCCCC 670
Db 421 GGCCTGGCGGCGAGAGCACACCTGCTGTCACAGGAGCACAGGAGCATGAAGACCCCC 480
QY 671 GTGAGTGTGGCGTCAAGTGGGATGACACCTCGGCTTCAAGCCGCTGCCGAGGTGGC 730
Db 481 GTGAGTGTGGCGTCAAGTGGGATGACACCTCGGCTTCAAGCCGCTGCCGAGGTGGC 540
QY 731 TACCGGGTCAAGGCGAGGAGCATCGATGATGATGAGACTCTGTTTGGCAGCCACAGCAGC 790
Db 541 TACCGGGTCAAGGCGAGGAGCATCGATGATGATGAGACTCTGTTTGGCAGCCACAGCAGC 600
QY 791 ACCCGGCTTACCCACCCGACTTCGATCCGCGCTTGGGTGGAGAAAGGTAAACAGAACCA 850
Db 601 ACCCGGCTTACCCACCCGACTTCGATCCGCGCTTGGGTGGAGAAAGGTAAACAGAACCA 660
QY 851 GCGGTGGCGAGGAGGAGCATCGAAGCCCTTGGGGCAAGAGGAGCTGTGAGACCAACCCCC 910
Db 661 GCGGTGGCGAGGAGGAGCATCGAAGCCCTTGGGGCAAGAGGAGCTGTGAGACCAACCCCC 720
QY 911 TCAAGGGCGAGCACCCCGCCCTCACACCAAGGAGAGAAACAATAACAGACCCATCAGC 970
Db 721 TCAAGGGCGAGCACCCCGCCCTCACACCAAGGAGAGAAACAATAACAGACCCATCAGC 780
QY 971 CACACCCGCTTACTGTGATGAGTGGCTGTTGGCTCCCGATCTGAGGGCGGAGCTTC 1030
```



```
QY 1321 TGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGTGACCTCCGGAGCCCTCGTGAC 1380
Db 1321 TGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGTGACCTCCGGAGCCCTCGTGAC 1380
QY 1381 TTCCAGGCTCGCTCAGTTAGCATTTTCAGTGCCATCTACCCACGACGAGGTGGGCCAC 1440
Db 1381 TTCCAGGCTCGCTCAGTTAGCATTTTCAGTGCCATCTACCCACGACGAGGTGGGCCAC 1440
QY 1441 CCAGAACCAAGCCCTTGGAAATGATCTCTTTTCATCAGGGTTGCCATGGGCCAC 1500
Db 1441 CCAGAACCAAGCCCTTGGAAATGATCTCTTTTCATCAGGGTTGCCATGGGCCAC 1500
QY 1501 GGGCACAGGTATGGCCCTTGGCAGGATAGGACATTCATCACCAGGAGAACCCAGG 1560
Db 1501 GGGCACAGGTATGGCCCTTGGCAGGATAGGACATTCATCACCAGGAGAACCCAGG 1560
QY 1561 TATTAAAGAGCCCTTGTGGGGGAGACAGACATAGAGGGGTGGGAGTGCTTCCCTTT 1620
Db 1561 TATTAAAGAGCCCTTGTGGGGGAGACAGACATAGAGGGGTGGGAGTGCTTCCCTTT 1620
QY 1621 ATCTGACAACTCTAGTGCATTTCTGCTTTTCTCCGATGCGGATTTGGGGGCCAC 1680
Db 1621 ATCTGACAACTCTAGTGCATTTCTGCTTTTCTCCGATGCGGATTTGGGGGCCAC 1680
QY 1681 CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACATATCTCAAAATTAGTGCCAAACCCAGG 1740
Db 1681 CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACATATCTCAAAATTAGTGCCAAACCCAGG 1740
QY 1741 GGCTGACCTCTCCACATCATCTCTCTCTCCAGTGGGATTAAGTGGGCTGAT 1800
Db 1741 GGCTGACCTCTCCACATCATCTCTCTCTCCAGTGGGATTAAGTGGGCTGAT 1800
QY 1801 TGCCAACTCG 1811
Db 1801 TGCCAACTCG 1811

RESULT 3
AY007158 1839 bp mRNA linear PRI 31-AUG-2000
LOCUS Homo sapiens clone CDABP0113 mRNA sequence.
DEFINITION
ACCESSION AY007158
VERSION AY007158.1 GI:9956071
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Anderson, B., Wentland, M.A., Ricalente, J.Y., Liu, W. and Gibbs, R.A.
A 'double adaptor' method for improved shotgun library construction
Anal. Biochem. 236 (1), 107-113 (1996)
96207227
PUBMED 8619474
REFERENCE 2 (bases 1 to 1839)
AUTHORS Yu, W., Anderson, B., Worley, K.C., Muzny, D.M., Ding, Y., Liu, W.,
Ricalente, J.Y., Wentland, M.A., Lennon, G. and Gibbs, R.A.
Large-scale concatenation cDNA sequencing
Genome Res. 7 (4), 353-358 (1997)
97264341
PUBMED 9110174
REFERENCE 3 (bases 1 to 1839)
AUTHORS Zhou, J., Yu, W., Tang, H., Mei, G., Tsang, Y.T.M., Bouck, J., Gibbs, R.A.
and Margolin, J.F.
Direct Submission
Submitted (24-JUL-2000) Human Genome Sequencing Center and Texas
Children's Cancer Center, Baylor College of Medicine, Houston, TX
77030, USA
The clone request should be directed to Dr. J. Margolin at
Pediatrics-Hematology & Oncology, Texas Children's REIGIN Center
102514, Houston, Texas 77030, USA. Telephone: 713-770-4583
margolin@bcm.tmc.edu.
FEATURES Location/Qualifiers
```

```
source 1. .1839
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CDABP0113"
/dev stage="infant"
/note="from patient with acute lymphoblastic leukemia"

ORIGIN
Query Match 90.1%; Score 1631.4; DB 9; Length 1839;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1632; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 179 CCTGTTCTCGGTGCAAAAGTCTGGGTTCCTGGGTTCTGGATTTCGGGGCCGTTACAC 238
Db 1 CCTGTTCTCGGTGCAAAAGTCTGGGTTCCTGGGTTCTGGATTTCGGGGCCGTTACAC 60
QY 239 GTAGCCTGTGCGGCTCTCTCGGTTAGTTCGTCCGCGGGGTGCCCGGACCGCCTAG 298
Db 61 GTAGCCTGTGCGGCTCTCTCGGTTAGTTCGTCCGCGGGGTGCCCGGACCGCCTAG 120
QY 299 GCTGCGGGGTTCGGGGCCCGAGGATTCGGGCTGCAGATTGACGGGATCCCGGATG 358
Db 121 GCTGCGGGGTTCGGGGCCCGAGGATTCGGGCTGCAGATTGACGGGATCCCGGATG 180
QY 359 CACGCGCGCCCCCGCGCCCTCACCGAGGGTCCAGACCTGTGTGGGAAGAGGTCCGGGG 418
Db 181 CACGCGCGCCCCCGCGCCCTCACCGAGGGTCCAGACCTGTGTGGGAAGAGGTCCGGGG 240
QY 419 ACGGTCCTCAGAGATCCCGATGCTAGCAGCAGAGTGTCTCAGCTTTATAGGTGTGACC 478
Db 241 ACGGTCCTCAGAGATCCCGATGCTAGCAGCAGAGTGTCTCAGCTTTATAGGTGTGACC 300
QY 479 TACACATGTGACTTCACCTCAGTTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTT 538
Db 301 TACACATGTGACTTCACCTCAGTTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTT 360
QY 539 CACAGTGTGTGAGAGATTAAATGAACAAATGCTTGTAAAGTCTTTTGAGAGGGAG 598
Db 361 CACAGTGTGTGAGAGATTAAATGAACAAATGCTTGTAAAGTCTTTTGAGAGGGAG 420
QY 599 CCTCGAAGCAGGCGCTGGCGGAGACACCTGTCTCACCAGGACCCACAGGAGC 658
Db 421 CCTCGAAGCAGGCGCTGGCGGAGACACCTGTCTCACCAGGACCCACAGGAGC 480
QY 659 ATGAAGACCCCGCTGGAGCTGCGGCTCAGTGGGATGCAGACCTCTGGGCTTCAGCACCGC 718
Db 481 ATGAAGACCCCGCTGGAGCTGCGGCTCAGTGGGATGCAGACCTCTGGGCTTCAGCACCGC 540
QY 719 TGCCGAGTGGTACCGGCTCAAGCCAGGACGTCATATGTGATGAGACTCTGTTGGC 778
Db 541 TGCCGAGTGGTACCGGCTCAAGCCAGGACGTCATATGTGATGAGACTCTGTTGGC 600
QY 779 AGCCAGCAGGACACCCGCGCTACCCACCGGACTTCGATCCGCGCTTGGGTGGAGAGGCT 838
Db 601 AGCCAGCAGGACACCCGCGCTACCCACCGGACTTCGATCCGCGCTTGGGTGGAGAGGCT 660
QY 839 AACAGAACAGAGGCTGGGCAAGAGGATCGAAGGCTTGGGGGCAAGAGGAGCTGT 898
Db 661 AACAGAACAGAGGCTGGGCAAGAGGATCGAAGGCTTGGGGGCAAGAGGAGCTGT 720
QY 899 GAGACACCCCTCAAGGGGAGCAGACCCCGCTCACACCAAGGAAGAAACAATATAC 958
Db 721 GAGACACCCCTCAAGGGGAGCAGACCCCGCTCACACCAAGGAAGAAACAATATAC 780
QY 959 AGACCCATCAGCCACACCCCGTCTTCTGTGATGAGTGCCTGTTGGCTCCCGATCTGAA 1018
Db 781 AGACCCATCAGCCACACCCCGTCTTCTGTGATGAGTGCCTGTTGGCTCCCGATCTGAA 840
QY 1019 GCGCGACCTTCGGGGCCCGCGGATGCGGAGGGGATGCCGAAAGCTCCGTCCTCTC 1078
Db 841 GCGCGACCTTCGGGGCCCGCGGATGCGGAGGGGATGCCGAAAGCTCCGTCCTCTC 900
```

```

RESULT 2
AK075358
LOCUS       AK075358               1811 bp      mRNA      linear      PRI 03-SEP-2002
DEFINITION  Homo sapiens cDNA P8EC0043 fis, clone NT2RP1001002.
ACCESSION  AK075358
VERSION    AK075358.1 GI:22761393
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
AUTHORS    Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayaishi, K.,
            Ishii, S., Sato, K., Yamamoto, J., Wakamatsu, A., Negai, T.,
            Nakamura, Y., Nagahari, K., Sugano, S. and Isogai, T.
TITLE      HRI human cDNA sequencing project
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1811)
AUTHORS    Isogai, T. and Yamamoto, J.
TITLE      Direct Submission
JOURNAL    Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
            HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass
            sequencing, clone selection and full insert sequencing; Helix
            Research Institute (supported by Japan Key Technology Center etc.);
            cDNA library construction; Institute of Medical Science, University
            of Tokyo, Laboratory of Genome Structure, Human Genome Center.
FEATURES   Location/Qualifiers
            1..1811
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="NT2RP1001002"
             /cell_line="NT2"
             /cell_type="teratocarcinoma"
             /clone_lib="NT2RP1"
             /note="Cloning vector: pUC19FL3-mRNA from NT2 neuronal
             precursor cells after 48-hours retinoic acid (RA)
             induction"
             659..1468
             /note="unnamed protein product"
             /codon_start=1
             /protein_id="BAC11568.1"
             /db_xref="GI:22761394"
             /translation="WKTPVELAVSGMOTLGLQHRGGRVVKARTSYVDETLFGSPAG
             TRTPPPDPDPWVEKANRTRGVKEASKALGAKGSCETPSRGSTPLTPRKKNKXRP
             ISHTPSYCDSELSRSGASFGAPRVAKGDAKLRLWTPPTPRGSHSPRPAP
             LRAHPAGPSKTEPGPAADSKMSGLHSRPLKRLGSLHLNVPSTGHPSAP
             HTNGPQDLRPSTSGVTRFSLVTSRARSVSISVPSFTRPGATQKPKPPWK"

CDS
            100.08; Score 1811; DB 9; Length 1811;
            Best Local Similarity 100.08; Pred. No. 0;
            Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
Query Match      100.08; Score 1811; DB 9; Length 1811;
Best Local Similarity 100.08; Pred. No. 0;
Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTGCTCGGGCCCTCGGCGCCTAGTACACACGACCTGAGTGCGCAGGAGGAC 60
DB      1 AGTGCTCGGGCCCTCGGCGCCTAGTACACACGACCTGAGTGCGCAGGAGGAC 60

QY      61 CCTCTCCATGTTAGGACCTCTGGGCTCAGAGGCTGGCGCCCGCCCTGGCGGAC 120
DB      61 CCTCTCCATGTTAGGACCTCTGGGCTCAGAGGCTGGCGCCCGCCCTGGCGGAC 120

QY      121 TCCCGCCATCGCGGGCGCGAATGGTCCGGTCCGCTCCGAGTGGCTGGTGTCTCC 180
DB      121 TCCCGCCATCGCGGGCGCGAATGGTCCGGTCCGCTCCGAGTGGCTGGTGTCTCC 180

QY      181 TGGTGTGGGTGCAAGTGTCTGGGTCTGGATTCGAGTTCGCGGCGGCTTCAACGT 240
DB      181 TGGTGTGGGTGCAAGTGTCTGGGTCTGGATTCGAGTTCGCGGCGGCTTCAACGT 240

```

```

QY      241 AGCTGTGCGGCTCTCTCGGTGAGTCCGTCCGCGCGGCTCCCGCGGAGCGGCTAGGC 300
DB      241 AGCTGTGCGGCTCTCTCGGTGAGTCCGTCCGCGCGGCTCCCGCGGAGCGGCTAGGC 300

QY      301 TCGCGGGGTTCGGGGCCCGCAGGCATTCGGGCTGCAGATTCACGGGATCCCGGATGCA 360
DB      301 TCGCGGGGTTCGGGGCCCGCAGGCATTCGGGCTGCAGATTCACGGGATCCCGGATGCA 360

QY      361 CGCGCGCCCCCGCGCCCTTACCGAGCGGTGCAGACCTGGTGGGAAGAGTTCGGGAGC 420
DB      361 CGCGCGCCCCCGCGCCCTTACCGAGCGGTGCAGACCTGGTGGGAAGAGTTCGGGAGC 420

QY      421 GGTCTCTCAGATCCCGATGCTACGAGCCAAAGATGTCAGCTTTATAGGTGTGACCTA 480
DB      421 GGTCTCTCAGATCCCGATGCTACGAGCCAAAGATGTCAGCTTTATAGGTGTGACCTA 480

QY      481 CACATGTGACTTCACCTCAGTTTGTGATCCGTAATGGACAAATTCGAAGCTACTTCA 540
DB      481 CACATGTGACTTCACCTCAGTTTGTGATCCGTAATGGACAAATTCGAAGCTACTTCA 540

QY      541 CAGTCTGTGAGAGATTAATGAACAATGCTTGTAAAGCTCTTTCAGAGGAGGCC 600
DB      541 CAGTCTGTGAGAGATTAATGAACAATGCTTGTAAAGCTCTTTCAGAGGAGGCC 600

QY      601 TCGGAAGCAGGGCCCTGGCGCGCAGAGCACACCTGCTGTCCACGAGGACACACAGGAGCAT 660
DB      601 TCGGAAGCAGGGCCCTGGCGCGCAGAGCACACCTGCTGTCCACGAGGACACACAGGAGCAT 660

QY      661 GAAGACCCCGTGGAGCTGGCGGTGAGTGGGATGAGACCTTCGCGCTTCAGACCGGTG 720
DB      661 GAAGACCCCGTGGAGCTGGCGGTGAGTGGGATGAGACCTTCGCGCTTCAGACCGGTG 720

QY      721 CCGAGTGTCTACCGGTCAAGGCCAGGACGTCATATGTGATGAGACTCTGTGTCGAG 780
DB      721 CCGAGTGTCTACCGGTCAAGGCCAGGACGTCATATGTGATGAGACTCTGTGTCGAG 780

QY      781 CCGAGCAGGACCCCGGCTTACCCCAACCGACTTCGATCCGCCCTGGTGGAGAAAGCTAA 840
DB      781 CCGAGCAGGACCCCGGCTTACCCCAACCGACTTCGATCCGCCCTGGTGGAGAAAGCTAA 840

QY      841 CAGAACCCAGAGCGTGGCGCAGGAGGCATCGAGGCTTGGGGGCAAGAGGAGCGTGA 900
DB      841 CAGAACCCAGAGCGTGGCGCAGGAGGCATCGAGGCTTGGGGGCAAGAGGAGCGTGA 900

QY      901 GACCAACCCCTCAAGGGGCGAGCACCCCAACCTCACCAAGGAAGAAAGAAATACAG 960
DB      901 GACCAACCCCTCAAGGGGCGAGCACCCCAACCTCACCAAGGAAGAAAGAAATACAG 960

QY      961 ACCCATCAGCCACACCCCGTCTTACTGTGATGAGTGGCTGTTGGCTCCGATCTGAAG 1020
DB      961 ACCCATCAGCCACACCCCGTCTTACTGTGATGAGTGGCTGTTGGCTCCGATCTGAAG 1020

QY      1021 CGCAGTTCGGGGCCCGCGGATGGCAAGGGGATGCGCAAGCTCCGTCCTCTCTT 1080
DB      1021 CGCAGTTCGGGGCCCGCGGATGGCAAGGGGATGCGCAAGCTCCGTCCTCTCTT 1080

QY      1081 GTGGAGCCACACCACTTACCCCGAGGGGTAGCACTTCGCCCGCCCGCAGGAGGACCACT 1140
DB      1081 GTGGAGCCACACCACTTACCCCGAGGGGTAGCACTTCGCCCGCCCGCAGGAGGACCACT 1140

QY      1141 GCGAGCATTCACCCAGTGGTCCCTCAAGACAGAGCGGGGCGCAGCGGAGACTCCCA 1200
DB      1141 GCGAGCATTCACCCAGTGGTCCCTCAAGACAGAGCGGGGCGCAGCGGAGACTCCCA 1200

QY      1201 GAAGTATCTATGGTGGGTATACCTTCTCAGCCCTCTGAAGCGGGGACTTTCCTTC 1260
DB      1201 GAAGTATCTATGGTGGGTATACCTTCTCAGCCCTCTGAAGCGGGGACTTTCCTTC 1260

QY      1261 CCTCAGCCACCTGATGTCCCGACCACTGGTCTATCCAGCCACCACTGTCGCCCCCA 1320
DB      1261 CCTCAGCCACCTGATGTCCCGACCACTGGTCTATCCAGCCACCACTGTCGCCCCCA 1320

```

FEATURES		Helix Research Institute (JP)	
source	Location/Qualifiers		
	1..1811		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	659..1468		
CDS	/notes="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC32398.1"		
	/db_xref="GI:13121983"		
	/db_xref="RENTREMBL:CAC32398"		
	/translation="MKTPVELAVSGMOTLGLQHRCKGGVRRVKAISYVDETLFGSPAG TRETPDPFPWKEANRTGVGKEASKALGKSCETTPSRGSTPTTLTPRXKNYRP ISHTPSYDESLFGSGEASFCAPMAKDAKLRALLWTPPTPRGSHSRPREAP LRAIHPAGPSKTPRSGADSKLSMGLHSSRLPKRGLSHSLTNVSTGHPSATAP HTNGPDLRPSTSGVTFRSPPLVTSRARSVSISVPSPRREGGATQKPKPWK"		
ORIGIN			
Query Match		100.0%; Score 1811; DB 6; Length 1811;	
Best Local Similarity		100.0%; Pred. No. 0;	
Matches 1811; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	AGTGCCCTGCGGCCCTCGGGCCCTAGTACACACGACCTGAGTGAGTGGCCACGAGGAC	60
Db	1	AGTGCCCTGCGGCCCTCGGGCCCTAGTACACACGACCTGAGTGAGTGGCCACGAGGAC	60
Qy	61	CCTCTCCATGTTAGGACCTCTCGGCCCTCAGAGCGTGGCCGCCCTGGGCGGAC	120
Db	61	CCTCTCCATGTTAGGACCTCTCGGCCCTCAGAGCGTGGCCGCCCTGGGCGGAC	120
Qy	121	TCCCCCATCGGGGGCGGAATGGTCCGGTCCGCTCGCAGTGTCTGGTGGTCTCCC	180
Db	121	TCCCCCATCGGGGGCGGAATGGTCCGGTCCGCTCGCAGTGTCTGGTGGTCTCCC	180
Qy	181	TGGTTGCTGGGTGCAAGTGTCTGGGTTCTGGGTTCTGGATTCGCGGGCGGTTTCA	240
Db	181	TGGTTGCTGGGTGCAAGTGTCTGGGTTCTGGGTTCTGGATTCGCGGGCGGTTTCA	240
Qy	241	AGCTGTGCGGCTCTCGGGTGAATGCTCCGGGCGGCTGCGCGGACGCGCTAGGC	300
Db	241	AGCTGTGCGGCTCTCGGGTGAATGCTCCGGGCGGCTGCGCGGACGCGCTAGGC	300
Qy	301	TGCGGGGGTCCGGGGGCCAGGATTCGCGGGTGCAGATTGACGGGGATCCCGGATGCA	360
Db	301	TGCGGGGGTCCGGGGGCCAGGATTCGCGGGTGCAGATTGACGGGGATCCCGGATGCA	360
Qy	361	CCGCGCCCTCGGCGCCCTCAACGACGGGTCCAGACCTGGTGGGAAGAGTCCGGGAC	420
Db	361	CCGCGCCCTCGGCGCCCTCAACGACGGGTCCAGACCTGGTGGGAAGAGTCCGGGAC	420
Qy	421	GGGTCCCTGAGGATCCCGATGCTTACGAGCAAGATGCTCAGCTTTATAGGTGTGACCTA	480
Db	421	GGGTCCCTGAGGATCCCGATGCTTACGAGCAAGATGCTCAGCTTTATAGGTGTGACCTA	480
Qy	481	CACATGTGACTTCACTCAGTTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTTCA	540
Db	481	CACATGTGACTTCACTCAGTTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTTCA	540
Qy	541	CAGTGTCTGAGAGGATTAATGAACAATGCTTGTAAAGCTTTTGCAGGAGGAGCC	600
Db	541	CAGTGTCTGAGAGGATTAATGAACAATGCTTGTAAAGCTTTTGCAGGAGGAGCC	600
Qy	601	TCGGAAGCAGGCGCTCGCGCAGACACACCTGCTCTACACGGGACCAACAGGACGAT	660
Db	601	TCGGAAGCAGGCGCTCGCGCAGACACACCTGCTCTACACGGGACCAACAGGACGAT	660
Qy	661	GAAGACCCCGTGGAGCTGGCGGTCACTGGGATGACAGCCCTCGGCCCTCAGCACCGCTG	720
Db	661	GAAGACCCCGTGGAGCTGGCGGTCACTGGGATGACAGCCCTCGGCCCTCAGCACCGCTG	720
Qy	721	CCGAGGTGGCTACCGGGTCAAGGCCAGGACCTCATATGTGATGAGACTCTGTTGGCAG	780
Db	721	CCGAGGTGGCTACCGGGTCAAGGCCAGGACCTCATATGTGATGAGACTCTGTTGGCAG	780

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 03:23:33 ; Search time 7205 Seconds  
(without alignments)  
10894.411 Million cell updates/sec

Title: US-10-031-589-3

Perfect score: 1811

Sequence: 1 agtcctgcggccctcggcg.....cggcgtgatgcgaacctgg 1811

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.man.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1811	100.0	1811	6	AX077432	Sequence
2	1811	100.0	1811	6	AX075358	Homo sapi
3	1631.4	90.1	1839	9	AY007158	Homo sapi
4	1621	89.5	1775	6	AX880378	Sequence
5	1621	89.5	1775	6	BD158346	Primer fo
6	1621	89.5	1775	9	AK027733	Homo sapi
7	1597.6	88.2	1709	9	AK123762	Homo sapi
8	1434.8	79.2	1614	6	AX880429	Sequence
9	1434.8	79.2	1614	6	BD158376	Primer fo
10	1434.8	79.2	1614	6	AX027741	Homo sapi
11	1371	75.7	1407	9	BC022092	Sequence
12	854	47.2	19522	9	AC089899	Homo sapi
13	661.4	36.5	784	6	AX869030	Sequence
14	661.4	36.5	784	6	BD149092	Primer fo
15	562	31.0	597	6	AX869102	Sequence
16	562	31.0	597	6	BD149164	Primer fo
17	550	30.4	816	6	AX781130	Sequence
18	494	27.3	561	6	AX874097	Sequence
19	494	27.3	561	6	BD154159	Primer fo
20	472.8	26.1	1485	10	BC021365	Mus muscu
21	369	20.4	572	6	AX779761	Sequence
22	353.8	19.5	263827	2	AC094579	Rattus no
23	353.8	19.5	274026	2	AC129633	Rattus no
24	341.8	18.9	578	6	AX874033	Sequence
25	341.8	18.9	578	6	BD154095	Primer fo
26	299.2	16.5	237928	2	AC137208	Rattus no
27	273.2	15.1	221622	10	AC125183	Mus muscu
28	241	13.3	199165	2	AC131195	Rattus no
29	186.2	10.3	60319	2	AC102037	Mus muscu
30	130.8	7.2	292	6	AX315440	Sequence
31	126	7.0	60319	2	AC102037	Mus muscu
32	125	6.9	232481	2	AC121112	Mus muscu
33	82.8	4.6	125020	9	AF429315	Homo sapi
34	61.6	3.4	125020	9	AF429315	Homo sapi
35	54.4	3.0	39970	9	AP001167	Homo sapi
36	54.4	3.0	146451	2	AC019309	Homo sapi
37	54.4	3.0	185984	9	AP000942	Homo sapi
38	54.2	3.0	146574	2	AC023459	Homo sapi
39	54.2	3.0	168860	9	AL160175	Human DNA
40	54	3.0	2000	6	AX655393	Sequence
41	53.8	3.0	3116	9	HSAC001552	Homo sapi
42	53.8	3.0	84544	2	AC009012	Homo sapi
43	53.8	3.0	159420	2	AC009017	Homo sapi
44	53.8	3.0	160672	9	AC008608	Homo sapi
45	53.6	3.0	89992	9	AL135794	Human DNA

ALIGNMENTS

RESULT 1	AX077432	Sequence 3 from Patent WO0107607.	1811 bp	DNA	linear	PAT 22-FEB-2001
LOCUS	AX077432	Sequence 3 from Patent WO0107607.	1811 bp	DNA	linear	PAT 22-FEB-2001
DEFINITION	AX077432	Sequence 3 from Patent WO0107607.	1811 bp	DNA	linear	PAT 22-FEB-2001
ACCESSION	AX077432	Sequence 3 from Patent WO0107607.	1811 bp	DNA	linear	PAT 22-FEB-2001
VERSION	AX077432.1	GI:13121982	1811 bp	DNA	linear	PAT 22-FEB-2001
KEYWORDS	AX077432.1	GI:13121982	1811 bp	DNA	linear	PAT 22-FEB-2001
SOURCE	AX077432.1	GI:13121982	1811 bp	DNA	linear	PAT 22-FEB-2001
ORGANISM	AX077432.1	GI:13121982	1811 bp	DNA	linear	PAT 22-FEB-2001
REFERENCE	AX077432.1	GI:13121982	1811 bp	DNA	linear	PAT 22-FEB-2001
AUTHORS	AX077432.1	GI:13121982	1811 bp	DNA	linear	PAT 22-FEB-2001
TITLE	AX077432.1	GI:13121982	1811 bp	DNA	linear	PAT 22-FEB-2001
JOURNAL	AX077432.1	GI:13121982	1811 bp	DNA	linear	PAT 22-FEB-2001

This Page Blank (uspto)

Db 444 TCACTAAGACACAGTTTCTTTATCTGTAAATAGTGACATGATAGCATCTTGAAGGGC 385  
QY 547 TGTTCAGAGGATTAAATGAACAATGCTTGTAAAGCTCTTTGACAGGAGGAGCCTCGAA 606  
Db 384 TACTGTGAGGATTATGAAATATGACGCAATGTACTTAGCAGGTTCTGGACTCAAAA 325  
QY 607 GCAGGCGCTGCCCGCAGAG 626  
Db 324 AGGCAACTGTTATGATGAG 305

## RESULT 14

US-10-292-798-1243  
; Sequence 1243, Application US/10292798.  
; Publication No. US20030235833A1  
; GENERAL INFORMATION:  
; APPLICANT: SUWA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ABURATANI, HIROYUKI  
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
; FILE REFERENCE: 084335/166  
; CURRENT APPLICATION NUMBER: US/10/292,798  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 10/017,161  
; PRIOR FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: JP 2001-246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2070  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1243  
; LENGTH: 43981  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; LOCATION: source  
; FEATURE:  
; LOCATION: (1)...(43981)  
; NAME/KEY: CDS  
; LOCATION: (201)...(470)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (12636)...(12748)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (24946)...(25088)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (32801)...(32997)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (41097)...(41379)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (43585)...(43781)  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (2811)...(2910)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (6914)...(6914)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (6923)...(6923)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (6932)...(6932)  
; OTHER INFORMATION: a, t, c, g, unknown or other

; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (6956)...(6956)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (6959)...(6960)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (6964)...(6964)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (7905)...(7905)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; US-10-292-798-1243

Query Match 2.8%; Score 50.4; DB 16; Length 43981;  
Best Local Similarity 71.7%; Pred No. 0.00061;  
Matches 66; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 495 CTCACGTTTGTGATCCGTTAAATGACAAATTCGAAGCTACTTCCACAGTCTGTGTGAGA 554  
Db 18915 CCTCAGTTCTCTTATCTGTGTAGTGGAGATAATAACAACCTACCTCAAAGGGTTGTAGGA 18974  
QY 555 GGATTAAATGAACAATGCTTGTAAAGCTCTT 586  
Db 18975 GGATTAAATGAATTAATATTCGTAAGCACIT 19006

## RESULT 15

US-09-867-701-7274  
; Sequence 7274, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7274  
; LENGTH: 316  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; US-09-867-701-7274

Query Match 2.8%; Score 50; DB 9; Length 316;  
Best Local Similarity 67.0%; Pred No. 0.00015;  
Matches 71; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
QY 490 CTTACCTCAGTTTGTGATCCGTTAAATGGAAGCTACTTCCAGTCTGTGTG 549  
Db 7 CCTTACATCAATCTCCTTATCTGTGAAATGGGAATAATAAAATCTACCTCAGAAATTAT 66  
QY 550 TGAGAGGATTAAATGAACAATGCTTGTAAAGCTCTTTCGAGGAGG 595  
Db 67 TAACAGATTAAATGAGATAATCCATGGAAGCTAGTAGCAGCATG 112

Search completed: August 17, 2004, 09:56:45  
Job time : 860 secs

QY 550 TGACAGGATTAAATGAACAATGCTTGTAAAGCTCTTTGCAGGAGG 595  
| | | | | | | | | | | | | | | | | | | | | |  
Db 95 TAACAGAAATTAATGACATAATCCATGGAAGCTAGTAGCAGCATG 140

RESULT 11

US-10-242-355-240  
; Sequence 240, Application US/10242355  
; Publication No. US20030235831A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

Query Match	2.8%	Score 50.8	DB 16	Length 343
Best Local Similarity	66.0%	Pred. No. 9.1e-05		
Matches 70	Conservative	2	Mismatches 34	Indels 0
				Gaps 0

[illegible]

QY 550 TGAGAGGATTAAATGAACAATGCTTGTAAAGCTCTTTGCAGGAGG 595

Db 95 TAACAGAAATTAATGAGATAATCCATGGAAGTAGTAGCAGCATG 140

RESULT 12

```

1  RESUBMIT 12
2  US-10/027-632-102308/c
3  ; Sequence 102308, Application US/10027632
4  ; Publication No. US20020198371A1
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Wang, David G.
7  ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
8  ; Polymorphisms in the Human Genome
9  ; FILE REFERENCE: 108927, 129
10 ; CURRENT APPLICATION NUMBER: US/10/027,632
11 ; CURRENT FILING DATE: 2002-04-30
12 ; PRIOR APPLICATION NUMBER: US 60/218,006
13 ; PRIOR FILING DATE: 2000-07-12
14 ; PRIOR APPLICATION NUMBER: US 60/198,676
15 ; PRIOR FILING DATE: 2000-04-20

```

```

; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 102308
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-102308

```

## RESULT 13

```

US-10-027-632-102308/c
; Sequence 102308, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102308
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-102308

```

```

; Sequence 288030, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.123
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 288030
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-288030

Query Match      2.9%; Score 52.8; DB 16; Length 479;
Best Local Similarity 61.8%; Pred. No. 2.5e-05;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 474 TGACCTACACATGACTTCACCTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGC 533
DB 89 TGGCTGGGGAGGTAACCTTCACCTCAGTTTCTCATCTGAACATGGGTACAGTTGGCC 148

QY 534 TACTTCACAGTGTCTTGGAGAGGATTAAATGAACAATGTTGTAAAGCTCTTTTGCAGGA 593
DB 149 TGATTCAACAGTGTGTTGAGAGTAACTGACATGACGTGTAGAAAGCACCTTAGCAGAA 208

QY 594 GGGAGCCTCGGAAGCA 609
DB 209 TGCCTAGTCCAAAGCA 224

RESULT 9
US-10-242-355-1101
; Sequence 1101, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1102
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-1102

Query Match      2.8%; Score 51.6; DB 16; Length 680;
Best Local Similarity 67.9%; Pred. No. 6.5e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 490 CTTACCTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTACAGTGTGT 549
DB 35 CCTTACCTCAATCTCCTTATCTGTGAAATGGGAATAATAAAATCTACCTCAGAAATAT 94

QY 550 TGAGAGGATTAAATGAACAATGCTTTGAAAGCTTTTGCAGGAGG 595
DB 95 TAAACAAATTAATGAGATATCCATGGAAAGCTAGTAGCAGCATG 140

```

```

; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1101
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-1101

Query Match      2.8%; Score 51.6; DB 16; Length 680;
Best Local Similarity 67.9%; Pred. No. 6.5e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 490 CTTACCTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTACAGTGTGT 549
DB 35 CCTTACCTCAATCTCCTTATCTGTGAAATGGGAATAATAAAATCTACCTCAGAAATAT 94

QY 550 TGAGAGGATTAAATGAACAATGCTTTGAAAGCTTTTGCAGGAGG 595
DB 95 TAAACAAATTAATGAGATATCCATGGAAAGCTAGTAGCAGCATG 140

RESULT 10
US-10-242-355-1102
; Sequence 1102, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1102
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-1102

Query Match      2.8%; Score 51.6; DB 16; Length 680;
Best Local Similarity 67.9%; Pred. No. 6.5e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 490 CTTACCTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTACAGTGTGT 549
DB 35 CCTTACCTCAATCTCCTTATCTGTGAAATGGGAATAATAAAATCTACCTCAGAAATAT 94

```



```
Db 339 GTGGCTACCGGGTCAAGGCCAGACGTCATATGTGATGAGACTCTGTTTGGCAGCCAG 398
Qy 786 CAGGACCCCGGCTACCCACCGGACTTCGATCCGCCCTGGTGGAGAGGTTAAAGAA 845
Db 399 CAGGACCCCGGCTACCCACCGGACTTCGATCCGCCCTGGTGGAGAGGTTAAAGAA 458
Qy 846 CCAGAGGCGTGGCAGAGGAGGCATCG 871
Db 459 CCAGAGGCGTGGCAGAGGAGGCATCG 484

RESULT 5
US-09-864-408A-8425
; Sequence 8425, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No US20040009474A1 Human Polynucleotides and Polypeptides Enc
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8425
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-408A-8425

Query Match 7.2%; Score 130.8; DB 11; Length 292;
Best Local Similarity 72.8%; Pred. No. 4.9e-29;
Matches 182; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

Qy 639 CACGAGGGACCCAGGACGATGAGACCCCGCTGG-AGCTGGCCGCTCAGTGGGATGCG 697
Db 43 CACCACACCCACTGGGCAAGTGAGACCCCGCTGGCAGCCGCCAGCAGCGGATGAA 102
Qy 698 ACCTGGCGCTTCCAGCACCGCTGCCAGGTGGCTACCGGTTCAAGGCCAGGACGTCAT 757
Db 103 GCCATCCAGCTTCCAGCACCGCAGTCCACCAGTACAGGCTCAAGGCCAGGCGATCCTAT 162
Qy 758 GTGGATGAGACTCTGTTGGCAGCCAGCAGGACCCGCCCTACCCACCGGACTTCGAT 817
Db 163 GTGGATGAGACCTCTGTTGGCAGCCAGCAGGACCCAGCAGGACCCAGCAGCTTGC 222
Qy 818 CCGCCCTGGGTGAGAGGCTACAGAACCCAGGCGTGGGCAAGAGGCGATCGAAGGCC 877
Db 223 CCACCCCTGGGTGAGAGGCTTACAGGATGTTACCGATCCAGAGGAGTGGGCCAGGACCC 282
Qy 878 TTGGGGGCGA 887
Db 283 TCTCTGGCCA 292

RESULT 6
US-09-908-975-4892
; Sequence 4892, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; FILE REFERENCE: THAT POPULATE A TRANSCRIPTOME
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20

; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4892
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-4892

Query Match 3.3%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 8.2e-08;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 164 TGTGCTGGCTGCTCCCTGGTGTGCTGGTGCAAGTGCTGGGTTCTGGGTTCTGGATTC 223
Db 1 TGTGCTGGCTGCTCCCTGGTGTGCTGGTGCAAGTGCTGGGTTCTGGGTTCTGGATTC 60

RESULT 7
US-10-027-632-288030
; Sequence 288030, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 288030
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-288030

Query Match 2.9%; Score 52.8; DB 13; Length 479;
Best Local Similarity 61.8%; Pred. No. 2.5e-05;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 474 TGACCTACACATGACTTCACCTCAGTTTGTGATCCGTAAATGGACAAATTCGAAGC 533
Db 89 TGGCTGGGAGGTAACTTCACTCAGTGTCTCACTGACCATGGGTACAGTTGCC 148
Qy 534 TACTTCACAGTGTGTTGAGAGGATTAATGAACAATGCTTGAAGCTCTTTGAGGA 593
Db 149 TGATTCAAGGTGGTTGTAAAGATTAACATGACATGCGTGTGAAGAGCACTTAGCAGAA 208
Qy 594 GGGAGCCTCGAAGCA 609
Db 209 TGCCTAGTCCAAGCA 224

RESULT 8
US-10-027-632-288030
```

```

QY 922 CACCCACCCCTCACACCAAGGAAGAAACAATAACAGACCCATCAGCCACACACCCGTC 981
Db 254 CACCCACCCCTCACACCAAGGAAGAAACAATAACAGACCCATCAGCCACACACCCGTC 313
QY 982 TTAATGTGATGAGTCGCTGTTGGCTCCCGATCTGAAGCGCCAGCTTCGGGGCCCGCG 1041
Db 314 TTAATGTGATGAGTCGCTGTTGGCTCCCGATCTGAAGCGCCAGCTTCGGGGCCCGCG 373
QY 1042 GATGGCAAGGGGATGCGCCAAAGCTCCGTGCTCTCTTGAGACCCACCACTACCC 1101
Db 374 GATGGCAAGGGGATGCGCCAAAGCTCCGTGCTCTCTTGAGACCCACCACTACCC 433
QY 1102 CAGGGTAGCACTCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTCACCCAGCTGG 1161
Db 434 CAGGGTAGCACTCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTCACCCAGCTGG 493
QY 1162 TCCCTCCAGACAGAGCGGGCCAGCGGAGACTCCAGAGTTATCTATGGTGGTT 1221
Db 494 TCCCTCCAGACAGAGCGGGCCAGCGGAGACTCCAGAGTTATCTATGGTGGTT 553
QY 1222 AACTCTTACAGCGCCCTGAAGCGGGGACTTCCCATTCCTCACCACCTGAATGCC 1281
Db 554 AACTCTTACAGCGCCCTGAAGCGGGGACTTCCCATTCCTCACCACCTGAATGCC 613
QY 1282 CAGCACTGTGATCCAGCCACCAAGTGCCTCCCAACAATGGGCTCAGGATCTCAGGCC 1341
Db 614 CAGCACTGTGATCCAGCCACCAAGTGCCTCCCAACAATGGGCTCAGGATCTCANGCC 673
QY 1342 TTCACCTGAGGGTGACCTTCGGAGCCCGCTGGT 1377
Db 674 TTCACCTGAGGGTGACCTTCGGAGCCCGCTGGT 709

*
RESULT 3
US-09-918-995-1667
; Sequence 1667, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1667
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(477)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-1667

Query Match 24.1%; Score 436.6; DB 10; Length 477;
Best Local Similarity 99.1%; Pred. No. 1.3e-121;
Matches 439; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 986 TGTGATGAGTCGCTGTTGGCTCCCGATCTGAAGGCGCCAGCTTCGGGGCCCGCGATG 1045
Db 35 TGGAAATTCGTCGCTGTTGGCTCCCGATCTGAAGGCGCCAGCTTCGGGGCCCGCGATG 94
QY 1046 GCGAAGGGGATCGCCAAAGCTCCGTGCTCTCTTGAGACGCCACCACTACCCCAAG 1105
Db 95 GCGAAGGGGATCGCCAAAGCTCCGTGCTCTCTTGAGACGCCACCACTACCCCAAG 154
QY 1106 GGTAGCCACTCGCCCGCCAGGAGGACCACTGCGAGCCATTCACCCAGCTGTC 1165

```

```

Db 155 GGTAGCCACTCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTCACCCAGCTGGTCCC 214
QY 1166 TCCAAGACAGAGCGGGGCCAGCGGAGACTCCCAAGAAATTAATCTATGGTGGTTACAC 1225
Db 215 TCCAAGACAGAGCGGGGCCAGCGGAGACTCCCAAGAAATTAATCTATGGTGGTTACAC 274
QY 1226 TCTTACGCCCCCTGAAGCGGGGACTTCCCAATTCCTCACCACCTGAATGTCCCCAGC 1285
Db 275 TCTTACGCCCCCTGAAGCGGGGACTTCCCAATTCCTCACCACCTGAATGTCCCCAGC 334
QY 1286 ACTGTCATCCAGCACCAAGTGCCTCCCAACAATGGGCTCAGGATCTCAGGCTTCC 1345
Db 335 ACTGTCATCCAGCACCAAGTGCCTCCCAACAATGGGCTCAGGATCTCAGGCTTCC 394
QY 1346 ACCTCAGGGTGACCTTCGGAGCCCGCTGGTGAATCTTCAGGGGCTCGCTCAGTTAGCAT 1405
Db 395 ACCTCAGGGTGACCTTCGGAGCCCGCTGGTGAATCTTCAGGGGCTCGCTCAGTTAGCAT 454
QY 1406 TCAGTGCCATCTACCCACGAGC 1428
Db 455 TCAGTGCCATCTACCCACGAGC 477

RESULT 4
US-09-918-995-31166
; Sequence 31166, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31166
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(484)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31166

Query Match 22.7%; Score 410.6; DB 10; Length 484;
Best Local Similarity 93.9%; Pred. No. 9.5e-114;
Matches 419; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 426 CCTGAGGATCCCGATGCTTACGAGCCAGAGTCTCAGCTTTATAGGTGTGACCTACACAT 485
Db 39 CNGNAANNNNNNNAGNANACGACGAGTGTCTTATAGGTGTGACCTACTCAT 98
QY 486 GTGACTTCACCTCAGTTTGTGATCCCGTAAATTCGAAATTCGAAGCTACTTCACAGTG 545
Db 99 GTGACTTTACCTCAGTTTGTGATCCCAAAATTCGAAAGCTACTTCACAGAG 158
QY 546 CTGTTGAGAGATTAATGAACAATGCTTGAAGCTCTTTGAGGAGGAGCTCGGA 605
Db 159 CTGTTGAGAGATTAATGAACAATGCTTGAAGCTCTTTGAGGAGGAGCTCGGA 218
QY 606 AGCAGGGCTCGCGGAGAGACACACCTGCTCACCAGGACCCACAGGAGCATGAAGA 665
Db 219 AGCAGGGCTCGCGGAGAGACACACCTGCTCACCAGGACCCACAGGAGCATGAAGA 278
QY 666 CCCCCGTGAGCTGCGCTCAGTGGATGACAGACCTTCGGCTTCAGCAGCGCTGCCAG 725
Db 279 CCCCCGTGAGCTGCGCTCAGTGGATGACAGACCTTCGGCTTCAGCAGCGCTGCCAG 338
QY 726 GTGGCTACCGGCTCAGGCCAGGACGTCATATGTGGATGAGACTCTGTTGGCAGCCAG 785

```

Db 130 CCCAGGCAATTCGGGCTGCAGATTGACGGGGATCCGGATGCACCGCGGCCCCCGGC 189  
Qy 377 CTTACCGAGCGGTCCAGACTCGTGGGAAGAGGTGCGGGACGGTCCCTGAGATCC 436  
Db 190 CCTCACCGAGCGGTCCAGACTCGTGGGAAGAGGTGCGGGACGGTCCCTGAGATCC 249  
Qy 437 CGATGCTCAGACGACGATGCTCAGCTTTATAGGTGTGACCTTACACATGTGACTTCAC 496  
Db 250 CGATGCTCAGACGACGATGCTCAGCTTTATAGGTGTGACCTTACACATGTGACTTCAC 309  
Qy 497 TCAGTTTGTGATCCGTAAATCGACAATTCGAAGCTACTTACAGTCTGTGTGAGAG 556  
Db 310 TCAGTTTGTGATCCGTAAATCGACAATTCGAAGCTACTTACAGTCTGTGTGAGAG 369  
Qy 557 ATTAATGAACAATCTCTTGAAGCTCTTTCAGAGGAGGCTCGGAAGCAGGCGCTG 616  
Db 370 ATTAATGAACAATCTCTTGAAGCTCTTTCAGAGGAGGCTCGGAAGCAGGCGCTG 429  
Qy 617 GCCGGCAGACACACCTGCTGTCCACGAGGACACACAGGAGCATGAAGACCCCGTGGAG 676  
Db 430 GCCGGCAGACACACCTGCTGTCCACGAGGACACACAGGAGCATGAAGACCCCGTGGAG 489  
Qy 677 CTGGCGGTGAGTGGATGACACCTCGGCTTCAGCAGCGCTGCCAGGTGCTACCG 736  
Db 490 CTGGCGGTGAGTGGATGACACCTCGGCTTCAGCAGCGCTGCCAGGTGCTACCG 549  
Qy 737 GTCAAGGCCAGGACGTCATATGTGATGAGACTCTGTTGGCAGCCCGACGACCCCG 796  
Db 550 GTCAAGGCCAGGACGTCATATGTGATGAGACTCTGTTGGCAGCCCGACGACCCCG 609  
Qy 797 CTTACCCACCGGACTTCGATCCGCTCGGTTGGAGAGGTAAACAGAACCGAGCGGTG 856  
Db 610 CTTACCCACCGGACTTCGATCCGCTCGGTTGGAGAGGTAAACAGAACCGAGCGGTG 669  
Qy 857 GGCAAGGAGCATCGAAGGCTTGGGGCAAGAGGAGGTGTGAGACCAACCCCTCAAG 916  
Db 670 GGCAAGGAGCATCGAAGGCTTGGGGCAAGAGGAGGTGTGAGACCAACCCCTCAAG 729  
Qy 917 GGCAGCACCCCACTCCACCAAGGAAGAAACAATAACAGCCCAACGACACAC 976  
Db 730 GGCAGCACCCCACTCCACCAAGGAAGAAACAATAACAGCCCAACGACACAC 789  
Qy 977 CCGTCTTACTGATGATGCTGTTGGCTCCGATCTGAAGGCGGAGCTTCGGGCG 1036  
Db 790 CCGTCTTACTGATGATGCTGTTGGCTCCGATCTGAAGGCGGAGCTTCGGGCG 849  
Qy 1037 CCGCGGATGCGAAGGGGATGCGCAAAAGCTCCGTGCTCTCTTGTGGACGCAACACT 1096  
Db 850 CCGCGGATGCGAAGGGGATGCGCAAAAGCTCCGTGCTCTCTTGTGGACGCAACACT 909  
Qy 1097 ACCCCAGGGGTAGCACTCGCCCGCCAGGGAGCACTCGGAGCCATTCACCA 1156  
Db 910 ACCCCAGGGGTAGCACTCGCCCGCCAGGGAGCACTCGGAGCCATTCACCA 969  
Qy 1157 GCTGGTCCCTCCAGACAGAGCGGGCCAGCGGAGACTCCAGAGGTTATCTATGGT 1216  
Db 970 GCTGGTCCCTCCAGACAGAGCGGGCCAGCGGAGACTCCAGAGGTTATCTATGGT 1029  
Qy 1217 GGGTTACATCTTTCAGCGCCCTGAAAGCGGGACTTTCCCATTCCTCCACCACTGAT 1276  
Db 1030 GGGTTACATCTTTCAGCGCCCTGAAAGCGGGACTTTCCCATTCCTCCACCACTGAT 1089  
Qy 1277 GTCCCGAGCACTGCTATCCAGCCACAGTGTGCCCCCACAAATGGGCTCAGATCTC 1336  
Db 1090 GTCCCGAGCACTGCTATCCAGCCACAGTGTGCCCCCACAAATGGGCTCAGATCTC 1149  
Qy 1337 AGGCTTCCACGTCAGGGGTGACTTTCGGAGCCCCCTGTGTGACTTCCAGGGCTCGCTCA 1396  
Db 1150 AGGCTTCCACGTCAGGGGTGACTTTCGGAGCCCCCTGTGTGACTTCCAGGGCTCGCTCA 1209  
Qy 1397 GTTAGCATTCAGTGCACTACCCACAGCAGGTGGGGCCACCAAGAAACCAAGGCC 1456  
Db 1210 GTTAGCATTCAGTGCACTACCCACAGCAGGTGGGGCCACCAAGAAACCAAGGCC 1269

## RESULT 2

US-10-452-858C-78

; Sequence 78, Application US/10452858C

; Publication No. US20040086945A1

; GENERAL INFORMATION:

; APPLICANT: Sreekrishna, Kotikanyadanam

; APPLICANT: Gerwe, Gina S.

; APPLICANT: Toerner, Daniel R.

; TITLE OF INVENTION: HAIRLESS PROTEIN-INTERACTING PARTNER COMPLEXES AND METHODS THEREOF

; FILE REFERENCE: 8956P

; CURRENT APPLICATION NUMBER: US/10/452,858C

; NUMBER OF SEQ ID NOS: 94

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 78

; LENGTH: 709

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (21)..(707)

; OTHER INFORMATION:

; NAME/KEY: misc\_feature

; LOCATION: (670)..(695)

; OTHER INFORMATION: n represents a, c t or g

US-10-452-858C-78

Query Match 38.0%; Score 688.2; DB 17; Length 709;

Best Local Similarity 99.1%; Pred. No. 8.2e-198;

Matches 690; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 682 CGTCAGTGGGATCGAGACCTTCGGCTTCAGACCCGCTCGGAGGTGGCTACCGGTCAC 741

Db 14 CGTCGAGGGATGAGACCTTCGGCTTCAGACCCGCTCGGAGGTGGCTACCGGTCAC 73

Qy 742 GGCCAGGACGTCATATGTGATGAGACTCTGTTGGCAGCCCGACGAGGACCCGGCTAC 801

Db 74 GGCCAGGACGTCATATGTGATGAGACTCTGTTGGCAGCCCGACGAGGACCCGGCTAC 133

Qy 802 CCCACCGGACTTCGATCCGCTTCGGGTGGAGAGGCTTAAACAGAACGAGCGCTGGGCAA 861

Db 134 CCCACCGGACTTCGATCCGCTTCGGGTGGAGAGGCTTAAACAGAACGAGCGCTGGGCAA 193

Qy 862 GGAGGATCGAAGGCTTCGGGCGCAAGGGGAGCTGTGAGACCAACCCCTCAAGGGCAG 921

Db 194 GGAGGATCGAAGGCTTCGGGCGCAAGGGGAGCTGTGAGACCAACCCCTCAAGGGCAG 253





This page Blank (uspio)